

(12) DEMANDE INTERNATIONALE PUBLIÉE EN VERTU DU TRAITÉ DE COOPÉRATION
EN MATIÈRE DE BREVETS (PCT)

(19) Organisation Mondiale de la Propriété
Intellectuelle
Bureau international



(43) Date de la publication internationale
7 juin 2001 (07.06.2001)

PCT

(10) Numéro de publication internationale
WO 01/40497 A2

(51) Classification internationale des brevets⁷: C12P 19/34,
C12N 15/52, 15/63, 15/11, 9/00, C12Q 1/68, C07K 16/40,
G01N 33/53, C12N 15/10, 15/52, 9/00, C12Q 1/68

(21) Numéro de la demande internationale:
PCT/FR00/03311

(22) Date de dépôt international:
27 novembre 2000 (27.11.2000)

(25) Langue de dépôt: français

(26) Langue de publication: français

(30) Données relatives à la priorité:
99/15032 29 novembre 1999 (29.11.1999) FR
60/209,800 7 juin 2000 (07.06.2000) US

(71) Déposant (pour tous les États désignés sauf US): AVENTIS
PHARMA S.A. [FR/FR]; 20, avenue Raymond Aron,
F-92160 Antony (FR).

(72) Inventeurs; et

(75) Inventeurs/Déposants (pour US seulement): JEAN-
NIN, Pascale [FR/FR]; 52, rue Pierre Louvrier, F-92140
Clamart (FR); PERNODET, Jean-Luc [FR/FR]; 21,
rue des Jardins, F-94230 Cachan (FR); GUERINEAU,
Michel [FR/FR]; 79, boulevard Saint Marcel, F-75013
Paris (FR); SIMONET, Pascal [FR/FR]; 55, rue Pierre

Voyant, F-69100 Villeurbanne (FR). COURTOIS, Sophie
[FR/FR]; 165, rue de Paris, F-94220 Charenton le Pont
(FR). CAPPELLANO, Carmela [FR/FR]; 16, rue de
Neuilly, F-94120 Fontenay sous Bois (FR). FRANCOU,
François [FR/FR]; 76, boulevard de Lozère, F-91120
Palaiseau (FR). RAYNAL, Alain [FR/FR]; 52, avenue des
Tilleuls, F-91440 Bures sur Yvette (FR). BALL, Maria
[VE/VE]; Avenue Cardenal Quintera, Res. Cardenal
Quintero, Edif. 10, Piso 4, Apto 42, Merida, Edo.,
Merida (VE). SEZONOV, Guennadi [RU/FR]; 16, rue
Saint Sauveur, F-75002 Paris (FR). TUPHILE, Karine
[FR/FR]; 39/41, boulevard Dubreuil, F-91400 Orsay (FR).
FROSTEGARD, Asa [NO/NO]; Flatveit Skogsvei 7,
N-1450 Nesoddtangen (NO).

(74) Mandataire: BOUVET, Philippe; Aventis Pharma S.A.,
Direction Brevets, 20, avenue Raymond Aron, F-92165
Antony Cedex (FR).

(81) États désignés (national): AE, AG, AL, AM, AT, AU, AZ,
BA, BB, BG, BR, BY, BZ, CA, CH, CN, CR, CU, CZ, DE,
DK, DM, DZ, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU,
ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS,
LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO,
NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR,
TT, TZ, UA, UG, US, UZ, VN, YU, ZA, ZW.

(84) États désignés (régional): brevet ARIPO (GH, GM, KE,
LS, MW, MZ, SD, SL, SZ, TZ, UG, ZW), brevet eurasien
(AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), brevet européen

[Suite sur la page suivante]

(54) Title: METHOD FOR OBTAINING NUCLEIC ACIDS FROM AN ENVIRONMENT SAMPLE, RESULTING NUCLEIC
ACIDS AND USE IN SYNTHESIS OF NOVEL COMPOUNDS

(54) Titre: PROCÉDE D'OBTENTION D'ACIDES NUCLÉIQUES À PARTIR D'UN ÉCHANTILLON DE L'ENVIRONNE-
MENT, ACIDES NUCLÉIQUES AINSI OBTENUS ET LEUR APPLICATION À LA SYNTHÈSE DE NOUVEAUX COMPOSÉS

(57) Abstract: The invention concerns a method for preparing nucleic acids from an environment sample, more particularly a method
for obtaining a library of nucleic acids from a sample. The invention also concerns nucleic acids of nucleic acid libraries obtained
by said method their use in the synthesis of novel compounds, in particular novel compounds of therapeutic interest. The invent
further concerns novel means used in the method for obtaining said nucleic acids, such as novel vectors and novel processes for
preparing such vectors or recombinant host cells containing said nucleic acid. Finally, the invention concerns methods for detecting
a nucleic acid of interest within a library of nucleic acids resulting from said method, and nucleic acids detected by said method and
polypeptides encoded by said nucleic acids.

(57) Abrégé: La présente invention concerne un procédé de préparation d'acides nucléiques à partir d'un échantillon de l'environ-
nement, plus particulièrement un procédé d'obtention d'une collection d'acides nucléiques à partir d'un échantillon. L'invention
est également relative aux acides nucléiques ou aux collections d'acides nucléiques obtenus selon le procédé et leur application à
la synthèse de nouveaux composés, notamment de nouveaux composés d'intérêt thérapeutique. L'invention a également pour objet
les moyens nouveaux mis en oeuvre dans le procédé d'obtention d'acides nucléiques ci-dessus, tels que de nouveaux vecteurs et des
nouveaux procédés de préparation de tels vecteurs ou encore des cellules hôtes recombinantes comprenant un acide nucléique de
l'invention. L'invention concerne encore des procédés pour détecter un acide nucléique d'intérêt au sein d'une collection d'acides
nucléiques obtenus selon le procédé ci-dessus, ainsi que les acides nucléiques détectés par un tel procédé et les polypeptides codés
par de tels acides nucléiques.

WO 01/40497 A2



(AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, TR), brevet OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG).

En ce qui concerne les codes à deux lettres et autres abréviations, se référer aux "Notes explicatives relatives aux codes et abréviations" figurant au début de chaque numéro ordinaire de la Gazette du PCT.

Publiée:

- Sans rapport de recherche internationale, sera republiée dès réception de ce rapport.

Procédé d'obtention d'acides nucléiques à partir d'un échantillon de l'environnement, acides nucléiques ainsi obtenus et leur application à la synthèse de nouveaux composés

5 La présente invention concerne un procédé de préparation d'acides nucléiques à partir d'un échantillon de l'environnement, plus particulièrement un procédé d'obtention d'une collection d'acides nucléiques à partir d'un échantillon. L'invention est également relative aux acides nucléiques ou aux collections d'acides nucléiques obtenus
10 selon le procédé et leur application à la synthèse de nouveaux composés, notamment de nouveaux composés d'intérêt thérapeutique.

L'invention a également pour objet les moyens nouveaux mis en oeuvre dans le procédé d'obtention d'acides nucléiques ci-dessus, tels que de nouveaux vecteurs et des nouveaux procédés de préparation
15 de tels vecteurs ou encore des cellules hôtes recombinantes comprenant un acide nucléique de l'invention.

L'invention concerne encore des procédés pour détecter un acide nucléique d'intérêt au sein d'une collection d'acides nucléiques obtenus selon le procédé ci-dessus, ainsi que les acides nucléiques
20 détectés par un tel procédé et les polypeptides codés par de tels acides nucléiques.

L'invention a également trait à des acides nucléiques obtenus et détectés selon les procédés ci-dessus, en particulier des acides nucléiques codant pour une enzyme participant à la voie de biosynthèse
25 d'antibiotiques tels que les β -lactames, les aminoglycosides, les nucléotides hétérocycliques ou encore des polykétides ainsi que l'enzyme codée par ces acides nucléiques, les polykétides produits grâce à l'expression de ces acides nucléiques et enfin des compositions pharmaceutiques comprenant une quantité pharmacologiquement active
30 d'un polykétide produit grâce à l'expression de tels acides nucléiques.

Depuis la découverte de la production de la streptomycine par les actinomycètes, la recherche de nouveaux composés d'intérêt thérapeutique, et tout particulièrement de nouveaux antibiotiques, a eu recours de manière accrue à des méthodes de criblage des métabolites
35 produits par les micro-organismes du sol.

De telles méthodes consistent principalement à isoler les organismes de la microflore tellurique, à les cultiver sur des milieux nutritifs spécialement adaptés puis à détecter une activité pharmacologique dans les produits retrouvés dans les surnageants de culture ou dans les lysats cellulaires ayant, le cas échéant, subi au préalable une ou plusieurs étapes de séparation et/ou de purification.

Ainsi, les méthodes d'isolement et de culture in vitro des organismes constituant la microflore tellurique ont permis, à la date d'aujourd'hui, de caractériser environ 40.000 molécules, dont environ la moitié présente une activité biologique.

Des produits majeurs ont été caractérisés selon de telles méthodes de culture in vitro, tels que des antibiotiques (pénicilline, érythromycine, actinomycine, tétracycline, céphalosporine), des anticancéreux, des anticholestérolémiants ou encore des pesticides.

Les produits d'intérêt thérapeutique d'origine microbienne connus à ce jour proviennent majoritairement (environ 70%) du groupe des actinomycètes et plus particulièrement du genre *Streptomyces*. Toutefois, d'autres composés thérapeutiques, tels que les teicoplanines, la gentamycine et les spinosines, ont été isolés à partir de micro-organismes de genres plus difficiles à cultiver tels que *Micromonospora*, *Actinomadura*, *Actinoplanes*, *Nocardia*, *Streptosporangium*, *Kitasatosporia* ou encore *Saccharomonospora*.

Mais la pratique illustre le fait que la caractérisation de nouveaux produits naturels synthétisés par les organismes de la microflore du sol est restée limitée, en partie du fait que l'étape de culture in vitro aboutit le plus souvent à une sélection d'organismes déjà connus antérieurement.

Les méthodes de séparation et de culture in vitro des organismes telluriques en vue d'identifier de nouveaux composés d'intérêt présentent donc de nombreuses limites.

Chez les actinomycètes, par exemple, le taux de redécouverte d'antibiotiques déjà connus antérieurement est d'environ 99%. En effet, des techniques de microscopie en fluorescence ont permis de dénombrier plus de 10^{10} cellules bactériennes dans 1g de sol, alors que

seulement 0,1 à 1% de ces bactéries peuvent être isolées après ensemencement sur des milieux de culture.

A l'aide de techniques de cinétique de réassociation d'ADN, il a pu être montré qu'entre 12.000 et 18.000 espèces bactériennes peuvent être contenues dans 1g de sol, alors qu'à ce jour, seuls 5000 micro-organismes non eucaryotes ont été décrits, tout habitat confondu.

Des études d'écologie moléculaire ont permis d'amplifier et cloner de nombreuses séquences nouvelles d'ADNr 16S à partir d'ADN de l'environnement.

Les résultats de ces études ont conduit à tripler le nombre de divisions bactériennes caractérisées antérieurement.

A la date d'aujourd'hui, les bactéries sont subdivisées en 40 divisions, certaines d'entre elles n'étant constituées que par des bactéries ne pouvant être cultivées. Ces derniers résultats témoignent de l'ampleur de la biodiversité microbienne restée inexploitée à ce jour.

Des travaux récents ont tenté de surmonter les nombreux obstacles à l'accès à la biodiversité de la microflore du sol, dont notamment l'étape de culture in vitro préalable à l'isolement et la caractérisation de composés d'intérêt industriel, surtout d'intérêt thérapeutique.

Des méthodes ont ainsi été mises au point qui incluent une étape d'extraction de l'ADN des organismes telluriques, le cas échéant après un isolement préalable des organismes contenus dans les échantillons de sol.

L'ADN ainsi extrait, après lyse des cellules bactériennes sans étape préalable de culture in vitro, est cloné dans des vecteurs utilisés pour transférer des organismes hôtes, afin de constituer des banques d'ADN provenant de bactéries du sol.

Ces banques de clones recombinants sont utilisées pour détecter la présence de gènes codant pour des composés d'intérêt thérapeutique ou alternativement pour détecter la production de composés d'intérêt thérapeutique par ces clones recombinants.

Toutefois, les méthodes d'accès direct à l'ADN de la microflore du sol, décrites dans l'état de la technique présentent des inconvénients lors de la mise en oeuvre de chacune des étapes décrites ci-dessus, de

nature à affecter considérablement la quantité et la qualité du matériel génétique obtenu et exploitable.

L'état de la technique concernant chacune des étapes de construction de banques d'ADN provenant d'échantillons de sol est
5 détaillé ci-après, ainsi que les inconvénients techniques identifiés par le demandeur et qui ont été surmontés selon la présente invention.

10 **1. Etape d'extraction de l'ADN à partir d'un échantillon du sol.**

1.1 Extraction directe d'ADN de l'environnement.

Il s'agit pour l'essentiel d'un procédé mettant en oeuvre des
15 techniques d'extraction d'ADN réalisées directement sur l'échantillon dans l'environnement, le plus souvent après une lyse in situ préalable des organismes de l'échantillon.

De telles techniques ont été mises en oeuvre sur des
20 échantillons provenant de milieux aquatiques, que ce soit d'eau douce ou marine. Elles comprennent une première étape de concentration préalable des cellules présentes librement ou sous forme de particules, consistant en général en une filtration de grands volumes d'eau sur différents dispositifs de filtration, par exemple filtration classique sur membrane, filtration tangentielle ou rotationnelle ou encore ultrafiltration.

25 La taille des pores est comprise entre 0,22 et 0,45 mm et nécessite souvent une préfiltration dans le but d'éviter des colmatages dus au traitement de grands volumes.

Dans un second temps, les cellules récoltées sont lysées
30 directement sur les filtres dans des petits volumes de solutions, par traitement enzymatique et/ou chimique.

Cette technique est par exemple illustrée par les travaux de
STEIN et al. , 1996, Journal of Bacteriology, Vol.178 (3): 591-599 qui décrit le clonage de gènes codant pour de l'ADN ribosomal et pour un
facteur d'élongation de la transcription (EF 2) à partir d'*Archaeobactéries*
35 du plancton marin.

Des techniques d'extraction directe d'ADN à partir d'échantillons de sol ou de sédiment ont été également décrites, basées sur des protocoles de lyse physique, chimique ou enzymatique réalisée in situ.

5 Par exemple, le brevet US N°5,824,485 (Chromaxome Corporation) décrit une lyse chimique des bactéries directement sur l'échantillon prélevé par addition d'un tampon de lyse chaud à base d'isothiocyante de guanidium.

La demande Internationale n°WO 99/20.799 (WISCONSIN
10 ALUMNI RESEARCH FOUNDATION) décrit une étape de lyse des bactéries in situ à l'aide d'un tampon d'extraction contenant une protéase et du SDS.

D'autres techniques ont également été utilisées telles que la réalisation de plusieurs cycles de congélation-décongélation de
15 l'échantillon puis pressage de l'échantillon décongelé à haute pression. Ont été également utilisées des techniques de lyse des bactéries à l'aide d'une succession d'étapes de sonication, de chauffage par micro-ondes et de chocs thermiques (PICARD et al. (1992).

Toutefois, les techniques d'extraction directe d'ADN de l'état de
20 la technique décrites ci-dessus ont une efficacité très variable du point de vue quantitatif et qualitatif.

Ainsi, les traitements chimiques ou enzymatiques in situ de l'échantillon ont le désavantage de ne lyser que certaines catégories de micro-organismes du fait de la résistance sélective des différents micro-
25 organismes indigènes à l'étape de lyse en raison de leur morphologie hétérogène.

Ainsi, les bactéries à Gram-positif résistent à un traitement à chaud au détergent SDS alors que la quasi-totalité des cellules à Gram-négatif sont lysées .

30 En outre, certains des protocoles d'extraction directe décrits ci-dessus favorisent l'adsorption des acides nucléiques extraits sur les particules minérales de l'échantillon, réduisant ainsi significativement la quantité d'ADN accessible.

Par ailleurs, si certains protocoles de l'état de la technique
35 divulguent une étape de traitement mécanique pour lyser les micro-

organismes de l'échantillon prélevé, une telle étape de lyse mécanique est systématiquement effectuée en milieu liquide dans un tampon d'extraction, ce qui ne permet pas une bonne homogénéisation de l'échantillon de départ sous la forme de particules fines permettant une
5 accessibilité maximale à la diversité des organismes présents dans l'échantillon. Des essais de broyage ont également été effectués sur échantillon de sol brut à l'aide de billes de verre, mais la quantité d'ADN extrait était faible.

Il a été observé selon l'invention qu'une première étape de lyse
10 mécanique in situ en milieu liquide avait des effets négatifs sur la quantité d'ADN susceptible d'être extrait.

La quantité d'ADN directement utilisable pour le clonage dans des vecteurs recombinants est également tributaire des étapes de purification subséquentes à son extraction.

15 Dans l'état de la technique, l'ADN extrait est ensuite purifié, par exemple par l'utilisation de polyvinylpyrrolidone, par une précipitation en présence d'acétate d'ammonium ou de potassium, par des centrifugations sur gradient de chlorure de césium, ou encore des techniques chromatographiques, notamment sur support
20 d'hydroxyapatite, sur colonne échangeuse d'ions ou encore tamisage moléculaire ou par des techniques d'électrophorèse sur gel d'agarose.

Les techniques de purification d'ADN décrites antérieurement, surtout lorsque celles-ci sont combinées avec les techniques d'extraction d'ADN de l'environnement précitées, sont susceptibles de conduire à
25 une co-purification de l'ADN avec des composés inhibiteurs provenant de l'échantillon initial qui sont difficiles à éliminer.

La co-extraction de composés inhibiteurs avec l'ADN nécessite la multiplication du nombre d'étapes de purification ce qui conduit à des pertes importantes de l'ADN initialement extrait et réduit simultanément
30 la diversité du matériel génétique initialement contenu dans l'échantillon, ainsi que sa quantité.

Un autre but de l'invention a été de surmonter les inconvénients des protocoles de purification antérieurs et de mettre au point une étape de purification d'ADN permettant de maintenir de manière optimale la

diversité de l'ADN de l'échantillon initial, d'une part, et, de favoriser quantitativement son obtention, d'autre part.

- Tout particulièrement, les améliorations qualitatives et quantitatives à la purification d'ADN sont maximales lorsqu'elles font appel à une combinaison d'un procédé d'extraction direct de l'ADN selon l'invention et d'un procédé de purification ultérieur, comme cela sera décrit ci-après.

1.2. Extraction indirecte d'ADN de l'environnement.

10

De telles techniques ont recours à une première étape de séparation des différents organismes de la microflore tellurique des autres constituants de l'échantillon de départ, préalablement à l'étape d'extraction de l'ADN proprement dite.

15

Dans l'état de la technique, la séparation préalable d'une fraction microbienne d'un échantillon de sol comprend le plus souvent une dispersion physique de l'échantillon par broyage de ce dernier en milieu liquide, par exemple en utilisant des dispositifs du type Waring Blender ou encore un mortier.

20

Il a également été décrit des dispersions chimiques, par exemple sur des résines échangeuses d'ions ou encore des dispersions à l'aide de détergents non spécifiques tels que le déoxycholate de sodium ou du polyéthylène glycol. Quel que soit le mode de dispersion, l'échantillon solide doit être mis en suspension dans de l'eau, du tampon phosphate ou une solution saline.

25

L'étape de dispersion physique ou chimique peut être suivie d'une centrifugation sur gradient de densité permettant la séparation des cellules contenues dans l'échantillon et des particules de ce dernier, étant entendu que les bactéries ont des densités inférieures à celles de la plupart des particules du sol.

30

L'étape de dispersion physique peut aussi être suivie alternativement d'une étape de centrifugation à faible vitesse ou encore une étape d'élutriation cellulaire.

35

L'ADN peut ensuite être extrait des cellules séparées par toutes les méthodes de lyse disponibles et être purifié par de nombreuses

méthodes, y compris les méthodes de purification décrites au paragraphe 1.1 précédent. Notamment, l'inclusion des cellules dans de l'agarose à bas point de fusion peut être réalisée afin de ménager la lyse.

5 Toutefois, les méthodes décrites dans l'état de la technique connues du demandeur ne donnent pas satisfaction du fait de la présence, dans les fractions contenant l'ADN extrait, de constituants indésirables de l'échantillon de départ ayant une influence significative sur la qualité et la quantité d'ADN final.

10 La présente invention se propose de résoudre les difficultés techniques rencontrées dans les procédés de l'art antérieur comme cela sera décrit ci-après.

2. Caractérisation moléculaire de l'ADN extrait.

15 Lorsque l'on désire construire une banque d'ADN à partir d'un échantillon de l'environnement, en particulier à partir d'un échantillon de sol, il est avantageux de vérifier la qualité et la diversité de la source d'ADN extrait et purifié préalablement à son insertion dans des vecteurs appropriés.

20 L'objectif d'une telle caractérisation moléculaire de l'ADN extrait et purifié est d'obtenir des profils représentant les proportions des différents taxons bactériens présents dans cet extrait d'ADN. La caractérisation moléculaire de l'ADN extrait et purifié permet de
25 déterminer si des artéfacts ont été introduits lors de la mise en oeuvre des différentes étapes d'extraction et de purification et, le cas échéant, si la diversité d'origine de l'ADN extrait et purifié est représentative de la diversité microbienne présente initialement dans l'échantillon, notamment dans l'échantillon de sol.

30 A la connaissance du demandeur, il est recouru dans l'état de la technique à des procédés d'hybridation quantitative mettant en oeuvre des sondes oligonucléotidiques spécifiques de différents groupes bactériens, appliqués directement à l'ADN extrait de l'environnement.

Malheureusement, une telle approche est peu sensible et ne permet pas de détecter des genres ou des groupes taxonomiques présents en faible abondance.

L'état de la technique décrit aussi des procédés de PCR quantitative, telle que la MPN-PCR ou encore la PCR quantitative par compétition. Toutefois, ces techniques présentent d'importants inconvénients.

Ainsi, la MPN-PCR est d'une utilisation complexe du fait de la multiplication des dilutions et des répétitions qui la rend inappropriée pour un grand nombre d'échantillons ou de couples d'amorces.

Par ailleurs, la PCR quantitative par compétition est d'une mise en oeuvre difficile du fait de la nécessité de construire un compétiteur spécifique à l'ADN cible qui, en outre, n'induit pas de biais ou d'artefacts dans la compétition proprement dite.

Il est ainsi proposé selon l'invention un procédé de précriblage d'une banque d'ADN provenant d'un échantillon de l'environnement qui est à la fois rapide, simple et fiable et permet de tester la qualité de l'ADN préalablement extrait et purifié et de déterminer ainsi l'intérêt de construire une banque de clones préparés à partir de cet ADN purifié de départ.

3. Vecteurs pour le clonage de l'ADN extrait et purifié à partir d'un échantillon de l'environnement.

De nombreux vecteurs ont déjà été décrits dans l'état de la technique afin de cloner de l'ADN préalablement extrait d'un échantillon de l'environnement.

Ainsi, selon la description de la demande internationale n°WO 99/20.799, peuvent être utilisés des vecteurs viraux, des phages, des plasmides, des phagemides, des cosmides, des phosmides, des vecteurs du type BAC (chromosome artificiel bactérien) ou encore le bactériophage P1, des vecteurs de type PAC (chromosome artificiel basé sur le bactériophage P1), des vecteurs du type YAC (chromosome artificiel de levure), des plasmides de levure ou tout autre vecteur

capable de maintenir et d'exprimer de manière stable un ADN génomique.

L'exemple 1 de la demande PCT n°WO 99/20.799 décrit la construction d'une banque d'ADN génomique par clonage dans un vecteur du type BAC.

A la connaissance du demandeur, aucune banque d'ADN provenant d'un échantillon de l'environnement n'avait encore été effectivement réalisée avec des vecteurs de type conjugatif, une telle technique étant rendue pour la première fois accessible et reproductible par l'homme du métier grâce à l'enseignement de la présente invention.

4. Hôtes cellulaires

Dans l'état de la technique, de nombreuses cellules hôtes ont été décrites comme pouvant être utilisées afin d'héberger les vecteurs contenant les inserts d'ADN provenant de l'ADN extrait et purifié à partir d'un échantillon de l'environnement.

Ainsi, la demande PCT N°WO 99/20.799 cite de nombreux hôtes cellulaires appropriés, tels que *Escherichia coli*, en particulier la souche DH 10B ou encore la souche 294 (ATCC 31446, la souche *E. coli* B, *E. Coli* X 1776 (ATCC N°31.537), *E.coli* DH5 α et *E.coli* W3110 (ATCC n°27.325).

Cette demande PCT cite également d'autres cellules hôtes appropriées telles que *Enterobacter*, *Erwinia*, *Klebsiella*, *Proteus*, *Salmonella*, *Serratia*, *Schigella* ou encore des souches du type bacillus telles que *B. subtilis* et *B. licheniformis* ainsi que les bactéries du genre *Pseudomonas*, *Streptomyces* ou *Actinomyces*.

Le brevet US N°5,824,485 cite en particulier la souche de *Streptomyces lividans* TK66 ou encore des cellules de levure telles que celles de *Saccharomyces pombe*.

5. Caractérisation de gènes d'intérêt dans des banques d'ADN provenant d'un échantillon de l'environnement

La demande PCT N° WO 99/20.799 décrit une identification du phénotype de différents clones appartenant à la banque d'ADN de *B.cereus*, respectivement un clone produisant de l'hémolysine, un clone hydrolysant l'esculine ou encore un clone produisant un pigment orange.

- 5 Des techniques de mutagenèse basées sur l'utilisation d'un transposon codant pour l'enzyme pho A ont permis subséquemment d'isoler des clones mutés et de caractériser les séquences responsables des phénotypes observés.

- 10 L'article de STEIN et al. (1996) précité décrit l'utilisation d'amorces spécifiques de l'ADN ribosomal afin d'amplifier l'ADN inséré dans les vecteurs hébergés par certains clones d'une banque d'ADN génomique d'Archaeobactéries de plancton marin et l'identification de plusieurs séquences codantes dans l'ADN ainsi amplifié.

- 15 L'article de BORSCHERT S. et al., (1992) décrit le criblage d'une banque d'ADN génomique de *Bacillus subtilis* à l'aide de couples d'amorces hybridant avec des régions conservées de peptide synthétases connues afin d'identifier un ou plusieurs gènes correspondant dans le génome de *Bacillus subtilis*.

- 20 Cette technique a permis de détecter un fragment d'ADN chromosomique d'environ 26 kb portant une partie de l'opéron de biosynthèse de la surfactine.

- 25 L'article de KAH-TONG S. et al.(1997) décrit le criblage d'une banque d'ADN provenant du sol à l'aide d'amorces hybridant avec des séquences conservées de l'opéron responsable de la voie de biosynthèse des polykétides de type II et montre l'identification, au sein de cette banque d'ADN, de séquences apparentées au gène PKS- β . Cet article décrit aussi la construction de cassettes d'expression hybrides dans lesquelles la séquence de la sous-unité PKS- β , retrouvée naturellement dans l'opéron responsable de la biosynthèse des polykétides, a été remplacée par différentes séquences apparentées retrouvées dans la banque d'ADN.

- 35 De même, l'article de HONG-FU et al. , (1995) décrit la construction de cassettes d'expression contenant les différentes phases de lecture ouverte de l'opéron responsable de la biosynthèse des polykétides, les différentes cassettes d'expression ayant été construites

artificiellement en combinant les phases de lecture ouverte qui ne sont pas retrouvées ensemble naturellement dans le génome de *Streptomyces coelicolor*. Cet article montre que la combinaison, dans les cassettes d'expression artificielles, de cadres de lecture ouvert
5 originaux de différentes souches bactériennes permet la production de polykétides ayant différentes caractéristiques structurales et des activités antibiotiques plus ou moins grandes vis-à-vis de *Bacillus subtilis* et *Bacillus cereus*.

Les polykétides font partie d'une grande famille de produits
10 naturels de structure variable et possédant une grande diversité d'activités biologiques. Font partie des polykétides par exemple, les tétracyclines et l'érythromycine (antibiotiques), le FK506 (immunosuppresseur), la doxorubicine (agent anti-cancéreux), la monensine (un agent coccidiostatique) ainsi que l'ivermectine (un agent antiparasitaire).
15

Ces molécules sont synthétisées grâce à des enzymes multifonctionnelles appelées polykétides synthases, qui catalysent des cycles de condensation répétés entre des acyl thioesters (en général des acétyl, propionyl, malonyl ou méthylmalonyl thioesters). Chaque cycle de
20 condensation aboutit à la formation, sur une chaîne croissante carbonée, d'un groupe β -céto qui peut ensuite subir, le cas échéant, une ou plusieurs séries d'étapes réductrices.

Compte-tenu de l'intérêt clinique important des polykétides, leur mécanisme commun de biosynthèse ainsi que le haut degré de
25 conservation observé entre les groupes de gènes codant pour les polykétides synthases, il s'est développé un intérêt accru pour le développement de nouveaux polykétides par génie génétique.

De nouveaux polykétides artificiels ont ainsi été produits par génie génétique, tels que la méderrhodine A ou la dihydrogranatirhodine.
30 La grande majorité des molécules nouvelles de polykétides obtenues par génie génétique sont très différentes, du point de vue structural, des polykétides correspondants naturels.

De l'état de la technique, il ressort ainsi qu'il existe un besoin d'obtention de nouveaux polykétides d'intérêt et tout particulièrement de
35 polykétides d'intérêt thérapeutique présentant notamment, par rapport à

leurs homologues naturels, un niveau accru d'activité antibiotique ou encore un spectre d'activité antibiotique différent, soit plus large que celui des polykétides connus, soit au contraire plus sélectif.

Ce besoin est, comme cela sera décrit ci-après, en partie
5 comblé selon la présente invention.

DESCRIPTION DE L'INVENTION

L'invention concerne tout d'abord un procédé pour la
10 construction de banques d'ADN provenant d'un échantillon de l'environnement, un tel échantillon pouvant être indifféremment un milieu aquatique (eau douce ou marine), un échantillon de sol (couche superficielle du sol, sous-sol ou sédiments), ou encore un échantillon d'organismes eucaryotes contenant une microflore associée, tel que par
15 exemple un échantillon provenant de plantes, d'insectes ou encore d'organismes marins et possédant une microflore associée.

La mise au point d'un procédé de construction d'une banque d'ADN d'un échantillon de l'environnement, et tout particulièrement d'un échantillon de sol, comprend des étapes critiques dont la mise en oeuvre
20 doit être nécessairement optimisée pour l'obtention d'une banque d'ADN dont le contenu en acides nucléiques d'intérêt répond aux objectifs initialement fixés.

Une première étape critique consiste en l'extraction et la purification ultérieure des acides nucléiques contenus initialement dans
25 l'échantillon, c'est-à-dire principalement des acides nucléiques contenus dans les divers organismes composant la microflore de cet échantillon.

La qualité de la purification de l'ADN extrait est déterminante sur le résultat obtenu.

Une seconde étape importante d'un procédé de construction
30 d'une banque d'acides nucléiques provenant d'un échantillon de l'environnement est l'évaluation de la diversité génétique des acides nucléiques extraits et purifiés. La mise au point d'une étape de réalisation simple et fiable de pré-criblage de l'ADN extrait et purifié afin de vérifier qu'il rend compte, au moins partiellement, de la diversité
35 phylogénétique des organismes présents initialement dans l'échantillon

de départ, permet en effet de déterminer l'intérêt ou non d'utiliser la source initiale d'ADN extrait et purifié pour la construction de la banque d'acides nucléiques proprement dite ou au contraire de ne pas poursuivre la construction de la banque d'acides nucléiques du fait d'artéfacts trop importants introduits au moment de l'extraction et de la purification des acides nucléiques. Il a en outre été identifié selon l'invention que la qualité des inserts introduits dans les vecteurs pour construire la banque est déterminante. Il a ainsi été déterminé que l'utilisation d'enzymes de restriction pour cliver l'ADN extrait et purifié à partir de l'échantillon de l'environnement était de nature à introduire des artéfacts ou "biais" dans la structure des inserts obtenus. En effet, l'ADN extrait du sol ou d'autres environnements, provenant en très grande majorité d'organismes non cultivables, est composé de molécules dont le taux de bases G et C est par définition inconnu et de plus variable en fonction de l'origine de ces organismes.

Une troisième étape critique est l'insertion des acides nucléiques extraits et purifiés dans des vecteurs capables d'intégrer des acides nucléiques de longueur choisie, d'une part, et, d'autre part, d'en permettre la transfection ou encore l'intégration dans le génome dans des hôtes cellulaires déterminés ainsi que, le cas échéant, d'en permettre l'expression dans de tels hôtes cellulaires.

Constituent des vecteurs d'intérêt, les vecteurs capables d'intégrer des acides nucléiques de grande taille, c'est-à-dire de taille supérieure à 100 kb lorsque l'objectif poursuivi consiste en un clonage et en une identification d'un opéron complet capable de diriger une voie complète de biosynthèse d'un composé d'intérêt industriel, en particulier d'un composé d'intérêt pharmaceutique ou agronomique.

DEFINITIONS

30

Au sens de la présente invention, on entend par "acides nucléiques", "polynucléotides" et "oligonucléotides" aussi bien des séquences d'ADN, d'ARN, que des séquences hybrides ARN/ADN de plus de 2 nucléotides, indifféremment sous la forme simple brin ou double brin.

35

Le terme " banque " ou " collection " est utilisé dans la présente description en référence indifféremment à un ensemble d'acides nucléiques extraits et, le cas échéant purifiés, provenant d'un échantillon de l'environnement, à un ensemble de vecteurs recombinants, chacun
5 des vecteurs recombinants de l'ensemble comprenant un acide nucléique provenant de l'ensemble d'acides nucléiques extraits et, le cas échéant purifiés précités, ainsi qu'à un ensemble de cellules hôtes recombinantes comprenant un ou plusieurs acides nucléiques provenant de l'ensemble des acides nucléiques extraits et, le cas échéant, purifiés
10 précités, lesdits acides nucléiques étant soient portés par un ou plusieurs vecteurs recombinants, soit intégrés dans le génome desdites cellules hôtes recombinantes.

On désigne par " échantillon de l'environnement " indifféremment un échantillon d'origine aquatique, par exemple d'eau
15 douce ou saline, ou un échantillon tellurique provenant de la couche superficielle d'un sol, de sédiments ou encore de couches inférieures du sol (sous-sol), ainsi que des échantillons d'organismes eucaryotes, le cas échéant multicellulaires, d'origine végétale, provenant d'organismes marins ou encore d'insectes et possédant une microflore associée, cette
20 microflore associée constituant des organismes d'intérêt.

On entend par " opéron " selon l'invention, un ensemble de cadres ouverts de lecture dont la transcription et/ou la traduction est co-régulée par un ensemble unique de signaux de régulation de la transcription et/ou de la traduction. Selon l'invention, un opéron peut
25 également comprendre lesdits signaux de régulation de la transcription et/ou de la traduction.

Par " voie métabolique " aux fins de l'invention ou encore " voie de biosynthèse " on entend un ensemble de réactions biochimiques anaboliques ou cataboliques réalisant la conversion d'une première
30 espèce chimique en une seconde espèce chimique.

Par exemple, une voie de biosynthèse d'un antibiotique est constituée de l'ensemble des réactions biochimiques convertissant des métabolites primaires en produits intermédiaires des antibiotiques, puis subséquentement en antibiotiques.

Par séquence de régulation placée "en phase" (en anglais operably linked) par rapport à une séquence nucléotidique dont l'expression est recherchée, on signifie que la ou les séquences de régulation de la transcription sont localisées, par rapport à la séquence
5 nucléotidique d'intérêt dont l'expression est recherchée, de manière à permettre l'expression de ladite séquence d'intérêt, la régulation de la dite expression étant dépendante de facteurs interagissant avec les séquences nucléotidiques régulatrices.

Selon une autre terminologie, on peut dire également que la
10 séquence nucléotidique d'intérêt dont l'expression est recherchée est placée "sous le contrôle" des séquences nucléotidiques régulatrices de la transcription.

Le terme "isolé" au sens de la présente invention désigne un matériel biologique qui a été soustrait à son environnement originel
15 (l'environnement dans lequel il est localisé naturellement).

Par exemple, un polynucléotide ou un polypeptide présent à l'état naturel dans un organisme (virus, bactérie, champignon, levure, plante ou animal) n'est pas isolé. Le même polypeptide séparé de son environnement naturel ou le même polynucléotide séparé des acides
20 nucléiques adjacents au sein desquels il est naturellement inséré dans le génome de l'organisme, est isolé.

Un tel polynucléotide peut être inclus dans un vecteur et/ou un tel polynucléotide peut être inclus dans une composition et demeure néanmoins à l'état isolé, du fait que le vecteur ou la composition ne
25 constitue pas son environnement naturel.

Le terme "purifié" ne nécessite pas que le matériel soit présent sous une forme de pureté absolue, exclusif de la présence d'autres composés. Il s'agit plutôt d'une définition relative.

Un polypeptide ou un polynucléotide est à l'état purifié après
30 purification du matériel de départ d'au moins un ordre de grandeur, de préférence 2 ou 3 et préférentiellement 4 ou 5 ordres de grandeur.

Le "pourcentage d'identité" entre deux séquences de nucléotides ou d'acides aminés, au sens de la présente invention, peut être déterminé en comparant deux séquences alignées de manière
35 optimale, à travers une fenêtre de comparaison.

La partie de la séquence nucléotidique ou polypeptide dans la fenêtre de comparaison peut ainsi comprendre des additions ou des délétions (par exemple des "gaps") par rapport à la séquence de référence (qui ne comprend pas ces additions ou ces délétions) de manière à obtenir un alignement optimal des deux séquences.

Le pourcentage est calculé en déterminant le nombre de positions auquel une base nucléique ou un résidu d'acide aminé identique est observé pour les deux séquences (nucléique ou peptidique) comparées, puis en divisant le nombre de positions auquel il y a identité entre les deux bases ou résidus d'acides aminés par le nombre total de positions dans la fenêtre de comparaison, puis en multipliant le résultat par 100 afin d'obtenir le pourcentage d'identité de séquence.

L'alignement optimal des séquences pour la comparaison peut être réalisé de manière informatique à l'aide d'algorithmes connus contenus dans le package de la Société WISCONSIN GENETICS SOFTWARE PACKAGE, GENETICS COMPUTER GROUP (GCG), 575 Science Doctor, Madison, WISCONSIN.

A titre d'illustration, le pourcentage d'identité de séquence pourra être effectué à l'aide du logiciel BLAST (versions BLAST 1.4.9 de Mars 1996, BLAST 2.0.4. de Février 1998 et BLAST 2.0.6. de Septembre 1998), en utilisant exclusivement les paramètres par défaut (S.F. Altschul et al., J. Mol. Biol. 1990 215: 403-410, S. F. Altschul et al., Nucleic Acids Res. 1997 25: 3389-3402). Blast recherche des séquences similaires/homologues à une séquence "requête" de référence, à l'aide de l'algorithme d'Altschul et al. La séquence requête et les bases de données utilisées peuvent être peptidiques ou nucléiques, toute combinaison étant possible.

EXTRACTION ET PURIFICATION D'ACIDES NUCLEIQUES PROVENANT D'UN ECHANTILLON DE L'ENVIRONNEMENT.

1. Extraction directe d'acides nucléiques

Il a été montré selon la présente invention que, pour l'obtention d'une banque d'acides nucléiques provenant d'organismes contenus

dans un échantillon du sol, il était important de créer des conditions dans lesquelles, d'une part, les différents organismes de l'échantillon sont rendus accessibles aux étapes ultérieures d'extraction des acides nucléiques et, d'autre part, que l'étape initiale de traitement de l'échantillon de sol permette une lyse mécanique maximale des organismes de l'échantillon de nature à rendre directement accessibles les acides nucléiques de ces organismes, principalement l'ADN génomique et plasmidique, aux tampons utilisés pour les étapes ultérieures d'extraction.

Il a été ainsi démontré selon l'invention qu'une accessibilité maximale des acides nucléiques provenant des micro-organismes d'un échantillon du sol était atteinte par un broyage poussé et à sec de l'échantillon de sol préalablement séché afin d'obtenir des micro-particules. Le demandeur a ainsi déterminé que le séchage de l'échantillon de sol préalable à tout traitement ultérieur provoque une diminution significative de la cohésion de l'échantillon de sol brut et favorise en conséquence sa désagrégation ultérieure sous la forme de micro-particules, lorsqu'un traitement par broyage approprié est opéré.

De manière surprenante, le demandeur a montré que des micro-particules d'échantillons de sol sec réunissaient des propriétés physico-chimiques favorables à l'extraction d'une quantité optimale d'acides nucléiques qui, dans leur nature, pouvaient être représentatifs de la diversité génétique des organismes présents initialement dans l'échantillon de sol de départ. Il a été montré en particulier que le procédé d'extraction directe d'acides nucléiques selon l'invention permettait l'extraction d'ADN provenant de micro-organismes rares, tels certains *Streptomyces* rares ou des micro-organismes sporulés.

Par "micro-particules" de l'échantillon de sol aux fins de la présente invention, on entend des particules dérivées de l'échantillon ayant une taille moyenne d'environ 50 μm , c'est à dire comprise en moyenne entre 45 et 55 μm .

Selon l'invention, les micro-particules sont obtenues à partir d'échantillons de sol préalablement séchés ou dessiqués puis broyés jusqu'à l'obtention de micro-particules de taille moyenne comprise entre

2µm et 50µm, avant remise en suspension dans un milieu tampon liquide des micro-particules obtenus.

Un tel milieu tampon liquide peut consister en un tampon d'extraction d'acides nucléiques, en particulier un tampon d'extraction d'ADN conventionnel bien connu de l'homme du métier.

Le broyage de l'échantillon de sol en micro-particules a pour double fonction de lyser mécaniquement la majorité des organismes présents dans l'échantillon de sol initial et de rendre accessibles les organismes non lysés par ce traitement mécanique à des étapes facultatives ultérieures de lyse chimique et/ou enzymatique.

Ainsi, un premier objet de l'invention consiste en un procédé de préparation d'une collection d'acides nucléiques à partir d'un échantillon de sol contenant des organismes, ledit procédé comprenant une première étape (I-(a)) d'obtention de micro-particules par broyage de l'échantillon de sol préalablement séché ou desséqué, puis mise en suspension des micro-particules dans un milieu tampon liquide.

De manière tout à fait préférée, l'étape de broyage est réalisée à l'aide d'un dispositif à billes d'agate ou de tungstène ou encore à l'aide d'un dispositif à anneaux de tungstène. Ces dispositifs sont préférés car la dureté de matériaux comme l'agate ou le tungstène facilite significativement l'obtention des micro-particules de la taille spécifiée ci-dessus. Pour cette raison, on ne choisira pas préférentiellement, voire on évitera, un recours à un dispositif de broyage à billes de verre, qui s'est révélé beaucoup moins efficace.

Le séchage ou la classification de l'échantillon de sol peut-être réalisée par toute méthode connue de l'homme du métier. Par exemple, l'échantillon de sol brut peut être séché à température ambiante pendant une durée de 24 à 48 heures.

Comme indiqué précédemment, le milieu tampon liquide peut consister en un milieu d'extraction de l'ADN présent dans les micro-particules. On utilisera de manière tout à fait préférée un tampon d'extraction désigné TENP contenant respectivement 50 mM tris, 20 mM EDTA, 100 mM NaCl et 1% (poids/volume) de polyvinylpyrrolidone, à pH 9,0.

Le procédé de préparation d'une collection d'acides nucléiques à partir d'un échantillon de sol est en outre caractérisé en ce que l'étape d'obtention de micro-particules par broyage de l'échantillon de sol préalablement séché ou dessiqué est suivie d'une étape I-(b) d'extraction des acides nucléiques présents dans les micro-particules.

Il est constant que l'extraction des acides nucléiques est accompagnée d'une co-extraction de composés et/ou de constituants du sol indésirables nécessitant la purification ultérieure des acides nucléiques extraits, une telle étape de purification ultérieure devant être à la fois suffisamment sélective pour permettre l'élimination des composés et/ou constituants du sol indésirables et d'un rendement suffisant pour entraîner une perte faible en quantité de l'ADN préalablement extrait.

Il a été montré selon l'invention qu'une étape de purification de l'ADN extrait des micro-particules de l'échantillon de sol répondant aux critères de sélectivité et de rendement définis ci-dessus, comprend un traitement de l'ADN extrait par une combinaison de deux étapes successives de chromatographie, respectivement une chromatographie sur tamis moléculaire et une chromatographie d'échange d'anions.

Selon une autre caractéristique du procédé ci-dessus, l'étape I-(b) d'extraction des acides nucléiques est suivie d'une étape I-(c) de purification des acides nucléiques extraits à l'aide des deux étapes de chromatographie suivantes:

- passage de la solution contenant les acides nucléiques sur un tamis moléculaire, puis récupération des fractions d'élution enrichies en acides nucléiques;

- passage des fractions d'élution enrichies en acides nucléiques sur un support de chromatographie d'échange d'anions, puis récupération des fractions d'élution contenant les acides nucléiques.

La nature et l'ordre des étapes de chromatographie ci-dessus sont essentiels à une bonne sélectivité et un excellent rendement de

l'étape de purification de l'ADN préalablement extrait des micro-particules de l'échantillon du sol préalablement séché ou dessiqué.

De manière très avantageuse, le support chromatographique du type " tamis moléculaire " de l'étape de purification d'acides nucléiques ci-dessus consiste en un support chromatographique de type Sephacryl®
5 S400 HR ou un support chromatographique de caractéristiques équivalentes.

De manière tout à fait préférée, le support chromatographique d'échange d'anions utilisé lors de la seconde étape de purification de
10 l'ADN extrait est un support de type Elutip® d, ou un support chromatographique de caractéristiques équivalentes.

En combinant les étapes I-(a) d'obtention de micro-particules de l'échantillon de sol sec, I-(b) d'extraction des acides nucléiques présents dans les micro-particules et I-(c) de purification par les étapes
15 chromatographiques décrites ci-dessus, il a été possible selon l'invention d'extraire directement l'ADN du sol sans purification préalable des cellules des organismes contenus initialement dans l'échantillon, tout en évitant la co-extraction de contaminants du sol, tels que par exemple les acides humiques qui est observée avec les procédés de l'état de la
20 technique.

Les contaminants, tels que les acides humiques affectent sévèrement les analyses et les utilisations subséquentes des acides nucléiques dont la purification est recherchée.

Selon le procédé ci-dessus, il est en outre possible d'accéder
25 aux acides nucléiques contenus dans les organismes qui n'ont pas été lysés mécaniquement au cours de l'étape I-(a) d'obtention de micro-particules de l'échantillon de sol, dans le but d'obtenir une collection quasi-exhaustive de la diversité génétique des acides nucléiques présents initialement dans l'échantillon de sol. Ainsi, les micro-particules
30 de l'échantillon de sol peuvent faire l'objet d'étapes ultérieures de traitement de lyse chimique, enzymatique ou physique, ou encore d'une combinaison de traitements chimiques, enzymatiques ou physiques.

Selon un premier aspect, le procédé de préparation d'une collection d'acides nucléiques à partir d'un échantillon de sol selon

l'invention, peut être en outre caractérisé en ce que l'étape I-(a) est suivie des étapes suivantes:

- traitement de la suspension de sol dans un milieu tampon
5 liquide par sonication;
- extraction et récupération des acides nucléiques.

De manière préférée, on aura recours, pour un traitement par
10 sonication, à un dispositif de type à micro-pointe en titane, tel que le dispositif 600 W Vibracell Ultrasonicator commercialisé par la Société Bioblock ou encore un sonicateur de type Cup Horn.

De manière tout à fait préférée, l'étape de sonication est
réalisée à une puissance de 15 W pendant une durée de 7 à 10 min et
15 comprend des cycles successifs de sonication, la sonication proprement dite étant réalisée pendant 50% de la durée de chaque cycle.

Selon un second aspect, le procédé ci-dessus peut être en
outre caractérisé en ce que l'étape I-(a) est suivie des étapes suivantes:

- 20 • traitement de la suspension de sol dans un milieu tampon liquide par sonication;
- incubation de la suspension à 37°C après sonication en présence de lysozyme et d'achromopeptidase;
- 25 • addition de SDS avant centrifugation et précipitation des acides nucléiques;
- récupération des acides nucléiques précipités.
- 30

De préférence, l'étape d'incubation en présence de lysozyme et d'achromopeptidase sera réalisée à une concentration finale de 0,3 mg/ml de chacune des deux enzymes, préférentiellement pendant 30 minutes à 37°C.

De manière préférée, le SDS sera utilisé à une concentration finale de 1% et pendant un temps d'incubation de 1 heure à la température de 60°C avant centrifugation et précipitation.

Selon un troisième aspect, le procédé de préparation d'une collection d'acides nucléiques à partir d'un échantillon de sol ci-dessus est en outre caractérisé en ce que l'étape I-(a) est suivie des étapes suivantes:

- homogénéisation de la suspension de sol avec une étape de mixage violent (vortex) suivie d'une étape de simple agitation;
- congélation de la suspension homogène suivie d'une décongélation ;
- traitement par sonication de la suspension après décongélation;
- incubation de la suspension à 37°C après sonication en présence de lysozyme et d'achromopeptidase;
- addition de SDS avant centrifugation et précipitation des acides nucléiques;
- récupération des acides nucléiques.

De manière préférée, les suspensions de micro-particules de sol sont passées au vortex puis homogénéisées par une agitation douce sur un agitateur à rotation circulaire pendant une durée de deux heures avant d'être congelées à -20°C.

Préférentiellement, les suspensions sont à nouveau agitées violemment par vortex pendant 10 minutes, après décongélation et avant l'étape de sonication.

Il va sans dire que les acides nucléiques extraits par les modes de réalisation du procédé d'extraction directe d'acides nucléiques décrit ci-dessus sont préférentiellement purifiés selon l'étape de purification constituée d'un premier passage sur tamis moléculaire puis un passage subséquent des fractions d'élution obtenues à l'issue de la chromatographie sur tamis moléculaire sur un support chromatographique d'échange d'anions.

2. Extraction indirecte des acides nucléiques

Selon un second mode de réalisation du procédé de préparation d'une collection d'acides nucléiques à partir d'un échantillon de l'environnement, selon l'invention, ledit échantillon de l'environnement subit un premier traitement de nature à permettre la séparation des organismes, contenus dans cet échantillon, des autres macro-constituants de l'échantillon.

Ce second mode de réalisation du procédé de préparation d'une collection d'acides nucléiques selon l'invention favorise l'obtention d'acides nucléiques de grande taille, qui sont pratiquement impossibles à obtenir selon le premier mode de réalisation du procédé selon l'invention décrit ci-dessus, l'étape de lyse mécanique opérée pour l'obtention des micro-particules ayant également pour effet de casser physiquement les acides nucléiques de l'échantillon de sol ou des acides nucléiques contenus dans les organismes de l'échantillon de sol.

L'obtention d'acides nucléiques de grande taille a été recherchée par le demandeur dans le but d'isoler et de caractériser les acides nucléiques comprenant, au moins partiellement, l'ensemble des séquences codantes appartenant à un même opéron capable de diriger la biosynthèse d'un composé d'intérêt industriel.

De manière préférée, on obtient, en mettant en oeuvre le second mode de réalisation du procédé de préparation d'une collection d'acides nucléiques à partir d'un échantillon de sol selon l'invention, des acides nucléiques ayant une taille supérieure à 100 kb, de préférence supérieure à 200, 250 ou 300 kb, et de manière tout à fait préférée d'acides nucléiques d'une taille supérieure à 400, 500 ou encore 600 kb.

Ce second mode de réalisation d'un procédé de préparation d'une collection d'acides nucléiques à partir d'un échantillon de l'environnement selon l'invention est constitué d'une combinaison de quatre étapes successives destinées à l'obtention des acides nucléiques ayant les caractéristiques décrites ci-dessus.

Lorsque l'échantillon de l'environnement est un échantillon de sol, il a été montré selon l'invention qu'une première étape d'obtention d'une suspension par dispersion de l'échantillon de sol en milieu liquide

favorisait l'accessibilité des organismes contenus dans l'échantillon sans provoquer de lyse mécanique significative des cellules.

La première étape d'obtention d'une dispersion de l'échantillon de sol ci-dessus rend accessibles les organismes de l'échantillon au milieu extérieur et permet également une dissociation partielle des organismes de l'échantillon et des macro-constituants. Elle rend ainsi possible une séparation ultérieure des organismes contenus initialement dans l'échantillon des autres constituants de ce dernier.

Lorsque l'échantillon de l'environnement provient par exemple de végétaux, d'organismes marins ou d'insectes, un traitement préalable par broyage est nécessaire afin de rendre les organismes de la microflore associée accessible aux étapes ultérieures du procédé.

Ainsi, le présent procédé comprend une étape de séparation des organismes des autres constituants minéraux et/ou organiques obtenus précédemment par une centrifugation sur un gradient de densité. Les organismes ainsi séparés sont ensuite soumis à une étape de lyse puis d'extraction des acides nucléiques.

L'étape de centrifugation sur un gradient de densité a, de manière surprenante, permis de séparer les cellules d'organismes des particules de sol contenues dans la suspension de l'échantillon. On aurait en effet pu s'attendre à ce qu'une proportion des cellules soient entraînées avec les macro-particules au sein de la phase de gradient. En outre, il n'avait jamais été démontré jusqu'à présent qu'une centrifugation sur gradient de densité d'un échantillon de sol permettait de retrouver, à l'interface phase aqueuse/gradient, une population d'organismes représentative de la diversité des organismes présents dans l'échantillon de départ, du fait que ces organismes sont de volume, densité et forme extrêmement variables. On pouvait raisonnablement supposer qu'ils seraient retrouvés indifféremment au sein de la phase aqueuse, à l'interface phase aqueuse/gradient de densité et également au sein du gradient de densité lui-même.

Ainsi, l'homme du métier pouvait s'attendre à ce que des organismes présentant des densités plus faibles ou plus grandes que la densité du gradient de densité utilisé (densité du gradient de densité

comprise entre 1,2 et 1,5 g/ml , préférentiellement 1,3 g/ml) ne pouvait être récupérés, ce qui aurait eu pour effet d'introduire un biais dans la représentativité des organismes effectivement séparés et, par voie de conséquence, également dans la diversité des acides nucléiques extraits.

En outre, dans un mode de réalisation particulier du procédé, une étape de germination des spores, en particulier d'actinomycètes, est réalisée, ce qui a pour effet d'accroître de manière significative la quantité d'ADN d'actinomycètes récupérée.

La dernière étape consiste en une étape de purification des acides nucléiques ainsi extraits sur un gradient de chlorure de césium.

De manière surprenante, la purification des acides nucléiques sur le gradient de chlorure de césium permet une élimination substantielle, voire complète, des substances composant le gradient de densité. Cette caractéristique est déterminante en ce qui concerne l'utilisation ultérieure des acides nucléiques purifiés car le gradient de densité est connu comme un puissant inhibiteur enzymatique, capable le cas échéant d'inhiber l'activité catalytique des enzymes utilisées pour préparer l'insertion des acides nucléiques extraits dans des vecteurs.

Selon ce second mode de réalisation, le procédé de préparation d'une collection d'acides nucléiques à partir d'un échantillon de l'environnement contenant des organismes selon l'invention comprend la succession d'étapes suivantes:

(i) obtention d'une suspension par dispersion de l'échantillon de l'environnement en milieu liquide puis homogénéisation de la suspension obtenue par agitation douce;

(ii) séparation des organismes des autres constituants minéraux et/ou organiques de la suspension homogène obtenue à l'étape (i) par centrifugation sur un gradient de densité;

(iii) lyse des microorganismes séparés à l'étape (ii) et extraction des acides nucléiques ;

(iv) purification des acides nucléiques sur un gradient de chlorure de césium .

Préférentiellement, la suspension de l'échantillon de sol est
5 obtenue par dispersion de cet échantillon par broyage à l'aide d'un
dispositif de type Waring Blender ou un dispositif de caractéristiques
équivalentes. De manière tout à fait préférée, la suspension d'échantillon
est obtenue après trois broyages successifs d'une durée d'une minute
chacun dans un dispositif de type Waring Blender. De préférence,
10 l'échantillon broyé sera refroidi dans la glace entre chacun des broyages.

De manière préférée, les organismes sont ensuite séparés des
particules du sol par centrifugation sur un coussin de densité du type
" Nycodenz ", commercialisé par la Société Nycomed Pharma AS. (Oslo
, Norvège). Les conditions préférées de centrifugation sont de 10.000g
15 pendant 40 minutes à 4°C, avantageusement dans un rotor à godets
mobiles du type " rotor TST 28.38 " commercialisé par la Société
KONTRON.

L'anneau d'organismes localisé, après centrifugation, à
l'interphase de la phase supérieure aqueuse et de la phase inférieure de
20 Nycodenz est alors prélevé et lavé par centrifugation avant reprise du
culot cellulaire dans un tampon approprié.

L'étape (iii) de lyse des organismes séparés à l'étape (ii) décrite
ci-dessus peut être réalisée de toute manière connue de l'homme du
métier.

25 Avantageusement, les cellules sont lysées dans une solution
Tris 10 mM-EDTA 100mM à pH 8.0 en présence de lysozyme et
d'achromopeptidase, avantageusement pendant une heure à 37°C.

L'extraction proprement dite de l'ADN peut être
avantageusement réalisée par addition d'une solution de lauryl sarcosyl
30 (1% du poids final de la solution) en présence de protéinase K et
incubation de la solution finale à 37°C pendant 30 minutes.

Les acides nucléiques extraits à l'étape (iii) sont ensuite purifiés
sur un gradient de chlorure de césium. Préférentiellement, l'étape de
purification des acides nucléiques sur un gradient de chlorure de césium

est réalisée par centrifugation à 35.000 tours/minute pendant 36 heures, par exemple sur un rotor du type Kontron 65.13.

Selon un aspect particulier du procédé de préparation d'une collection d'acides nucléiques à partir d'un échantillon de sol contenant
5 des organismes selon l'invention, lesdits acides nucléiques sont constitués majoritairement, sinon exclusivement, de molécules d'ADN.

Selon un autre aspect, les acides nucléiques peuvent être récupérés après inclusion des organismes, séparés sur un gradient de densité, dans un bloc d'agarose et lyse, par exemple chimique et/ou
10 enzymatique, des organismes inclus dans le bloc d'agarose.

Un autre objet de l'invention consiste en une collection d'acides nucléiques constitués des acides nucléiques obtenus à l'étape II-(iv) du procédé de préparation d'une collection d'acides nucléiques selon
15 l'invention ou encore obtenue à l'étape (c) ou une étape ultérieure du procédé de préparation d'une collection d'acides nucléiques selon l'invention.

L'invention est encore relative à un acide nucléique caractérisé en ce qu'il est contenu dans une collection d'acides nucléiques telle que
20 définie ci-dessus.

Selon un premier aspect, un tel acide nucléique constitutif d'une collection d'acides nucléiques selon l'invention est caractérisé en ce qu'il comprend une séquence nucléotidique codant au moins un opéron, ou une partie d'un opéron.

25 De manière tout à fait préférée, un tel opéron code pour la totalité ou une partie d'une voie métabolique.

L'exemple 9 décrit la construction d'une banque d'ADN génomique à partir d'une souche de *Streptomyces alboniger* et son clonage respectivement dans les cosmides navettes pOS700I et pOS700R. Il a été montré selon l'invention que dans la banque d'ADN
30 réalisée dans le vecteur intégratif pOS700I neuf clones contiennent des séquences nucléotidiques appartenant à l'opéron responsable de la voie de biosynthèse de la puromycine. De même, il a pu être identifié au sein de la banque d'ADN réalisée dans le vecteur réplcatif pOS 700R douze

clones contenant des séquences nucléotidiques de l'opéron responsable de la voie de biosynthèse de la puromycine.

En particulier, certains cosmides intégratifs et replicatifs des banques réalisées présentent, après digestion par les endonucléases de restriction ClaI et EcoRV, un fragment d'une taille de 12 kb susceptible de contenir la totalité des séquences de l'opéron responsable de la voie de biosynthèse de la puromycine.

Ainsi, selon un autre aspect, un acide nucléique selon l'invention contient, au moins en partie, des séquences nucléotidiques de l'opéron responsable de la voie de biosynthèse de la puromycine.

L'exemple 2 ci-après décrit la construction d'une banque d'ADN selon un procédé conforme à la présente invention dans un vecteur pBluescript SK⁺ à partir d'un sol contaminé par du lindane.

Les vecteurs recombinants ont été transfectés dans des cellules d'*Escherichia coli* DH10B puis les cellules transformées ont été cultivées dans un milieu de culture approprié en présence de lindane. Un criblage des clones de cellules transformées de la banque a permis de montrer que, sur 10.000 clones criblés, 35 d'entre eux présentaient un phénotype de dégradation du lindane. La présence du gène linA chez ces clones a pu être confirmée par amplification PCR grâce à des amorces spécifiques de ce gène.

Ainsi, selon un autre aspect, l'invention concerne également un acide nucléique contenant une séquence nucléotidique de la voie métabolique provoquant la biodégradation du lindane.

Il est donc clairement démontré, comme décrit plus haut, qu'un procédé de préparation d'une collection d'acides nucléiques à partir d'un échantillon de sol contenant des organismes selon l'invention ainsi qu'un procédé de préparation d'une collection de vecteurs recombinants contenant les acides nucléiques constitutifs de la collection d'acides nucléiques précités était tout à fait apte à l'isolement et à la caractérisation de séquences nucléotidiques incluses dans un opéron.

Une démonstration supplémentaire de l'aptitude d'un procédé selon l'invention à l'identification de séquences nucléotidiques codantes impliquées dans une voie de biosynthèse régulée sous la forme d'un opéron est en outre décrite plus loin: il s'agit du clonage et de la

caractérisation de séquences codant pour des polykétides synthases impliquées dans la voie de biosynthèse des polykétides, qui appartiennent à une famille de molécules dont certains représentants sont d'un intérêt thérapeutique majeur, en particulier antibiotique.

- 5 La présente invention a donc en outre pour objet un acide nucléique constitutif d'une collection d'acides nucléiques selon l'invention, caractérisé en ce qu'il comprend la totalité d'une séquence nucléotidique codant pour un polypeptide.

Selon un premier aspect, un acide nucléique constitutif d'une
10 collection d'acides nucléiques selon l'invention est d'origine procaryote.

Selon un second aspect, un acide nucléique constitutif d'une collection d'acides nucléiques selon l'invention provient d'une bactérie ou d'un virus.

Selon un troisième aspect, un acide nucléique constitutif d'une
15 collection d'acides nucléiques selon l'invention est d'origine eucaryote.

En particulier, un tel acide nucléique est caractérisé en ce qu'il provient d'un champignon, d'une levure, d'une plante ou d'un animal.

20 **CARACTERISATION MOLECULAIRE DE LA COLLECTION D'ACIDES
NUCLEIQUES EXTRAITS DU SOL.**

Afin de surmonter les nombreux inconvénients techniques des méthodes de caractérisation des banques d'ADN extraits et purifiés à partir d'un échantillon de l'environnement qui ont été décrits dans la
25 partie de la description relative à l'état de la technique, le demandeur a mis au point un procédé simple et fiable permettant de caractériser qualitativement et semi-quantitativement les acides nucléiques obtenus à l'issue du procédé décrit ci-dessus.

Le procédé selon l'invention consiste ainsi à amplifier
30 universellement un fragment de 700 pb localisé à l'intérieur d'une séquence d'ADN ribosomal de type 16 S, puis d'hybrider l'ADN amplifié avec une sonde oligonucléotidique de spécificité variable et enfin de comparer l'intensité d'hybridation de l'échantillon par rapport à une gamme étalon externe d'ADN de séquence ou d'origine connue.

L'amplification préalable à l'hybridation avec la sonde oligonucléotidique permet de quantifier des genres ou des espèces de micro-organismes peu abondants. De plus, l'amplification par des amorces universelles permet, lors de l'hybridation, d'utiliser une large

5 série de sondes oligonucléotidiques.

Ainsi, l'invention a en outre pour objet un procédé de détermination de la diversité des acides nucléiques contenus dans une collection d'acides nucléiques, et tout particulièrement d'une collection d'acides nucléiques provenant d'un échantillon de l'environnement,

10 préférentiellement d'un échantillon du sol, ledit procédé comprenant les étapes suivantes:

- mise en contact des acides nucléiques de la collection d'acides nucléiques à tester avec un couple d'amorces oligonucléotidiques hybridant à toute séquence d'ADN ribosomal 16 S

15 bactérien;

- réalisation d'au moins trois cycles d'amplification ;
- détection des acides nucléiques amplifiés à l'aide d'une sonde oligonucléotidique ou d'une pluralité de sondes oligonucléotidiques,

20 chaque sonde hybridant spécifiquement avec une séquence d'ADN ribosomal 16 S commune à un règne, un ordre, une sous-classe ou un genre bactérien;

- le cas échéant, comparaison des résultats de l'étape de détection précédente avec les résultats de détection, à l'aide de la sonde

25 ou de la pluralité de sondes d'acides nucléiques de séquence connue constituant une gamme étalon.

De manière préférée, un premier couple d'amorces hybridant avec des régions universellement conservées du gène de l'ARN ribosomal 16 S est constitué respectivement des amorces FGPS 612

30 (SEQ ID N°12) et FGPS 669 (SEQ ID N°13).

Un second mode de réalisation d'un couple d'amorces préféré selon l'invention est constitué du couple d'amorces universelles 63 f (SEQ ID N°22) et 1387r (SEQ ID N°23).

Selon un mode particulier de réalisation d'un procédé de

35 détermination de la diversité des acides nucléiques d'une collection

d'acides nucléiques, l'étape d'amplification à l'aide d'un couple d'amorces universelles peut être réalisée sur une collection de vecteurs recombinants dans chacun desquels a été inséré un acide nucléique de la collection d'acides nucléiques considérée, préalablement à l'étape
5 d'hybridation avec les sondes oligonucléotidiques spécifiques d'un règne, d'un ordre, d'une sous-classe ou d'un genre bactérien particulier.

Un tel procédé de détermination de la diversité des acides nucléiques contenus dans une collection est tout particulièrement applicable aux collections d'acides nucléiques obtenus conformément à
10 l'enseignement de la présente description.

Ainsi, l'exemple 3 détaille un procédé de préparation d'une collection d'acides nucléiques à partir d'un échantillon de sol contenant des organismes comprenant une étape d'extraction indirecte d'ADN par dispersion d'un échantillon du sol préalablement à la séparation des
15 cellules sur gradient de Nycodenz, lyse des cellules puis purification de l'ADN sur gradient de chlorure de césium.

La collection d'acides nucléiques ainsi obtenue a été utilisée telle quelle ou sous la forme d'inserts dans des vecteurs de type cosmide dans un procédé d'amplification à l'aide des amorces
20 universelles de l'ADNr 16 S précitées, puis les ADN amplifiés ont été soumis à une étape de détection à l'aide de sondes oligonucléotidiques de séquences SEQ ID N°14 à SEQ ID N°21 qui sont présentées dans le tableau 4.

Les résultats montrent qu'un procédé de préparation d'une
25 collection d'acides nucléiques à partir d'un échantillon de sol contenant des organismes selon l'invention permet d'accéder à l'ADN de plus de 14% de la microflore tellurique totale, soit 2×10^8 cellules par gramme de sol, alors que la microflore totale cultivable ne représente qu'à peine 2% de la population microbienne totale.

Afin de déterminer la diversité phylogénétique d'une collection
30 d'acides nucléiques préparés conformément à l'invention, 47 séquences du gène ARNr 16S ont été isolées et séquencées. Ces séquences correspondent respectivement aux séquences nucléotidiques SEQ ID N°60 à SEQ ID N°106.

Les acides nucléiques comprenant les séquences SEQ ID N° 60 à SEQ ID N° 106 font également partie de l'invention, ainsi que les acides nucléiques possédant au moins 99 %, préférentiellement 99,5% ou 99,8% d'identité en acides nucléiques avec les acides nucléiques comprenant les séquences SEQ ID N° 60 à SEQ ID N° 106. De telles séquences peuvent être utilisées notamment en tant que sondes pour cribler des clones d'une banque d'ADN et identifier ainsi ceux, parmi les clones de la banque, qui contiennent de telles séquences, ces séquences étant susceptibles d'être à proximité de séquences codantes d'intérêt, telles que des séquences codant pour des enzymes impliquées dans la voie de biosynthèse de métabolites antibiotiques, par exemple des polykétides.

La comparaison des séquences d'ARNr 16S à partir d'une banque d'ADN réalisée conformément à l'invention avec les séquences répertoriées dans la base données RDP (Maidak B.L., Cole J.R., Parker C.T., Garrity G.M., Larsen N., Li B., Lilburn T.G., McCaughey, M.J., Olsen G.J., Overbeek R., Pramanik S., Schmidt T.M., Tiedje J.M., Woese C.R. (1999) "A new projet of the RDP (Ribosomal Database Project)" Nucleic Acids Research Vol. 27: 171-173) ont permis de déterminer que les acides nucléiques contenus dans une collection d'acides nucléiques selon l'invention proviennent d' α -protéobactéries, de β -protéobactéries, de δ -protéobactéries, de γ -protéobactéries, d'actinomycètes ainsi que d'un genre apparenté à acidobactérium. Ces résultats, présentés dans le tableau 7 ainsi que par l'arbre phylogénétique de la figure 7 rendent compte de la grande diversité phylogénétique des acides nucléiques contenus dans une banque d'ADN préparée conformément au procédé selon l'invention.

VECTEURS DE CLONAGE ET/OU D'EXPRESSION

30

Chacun des acides nucléiques contenus dans une collection d'acides nucléiques préparés conformément à l'invention peut être inséré dans un vecteur de clonage et/ou d'expression.

A cette fin, tous types de vecteurs connus de l'état de la technique peuvent être utilisés, tels que des vecteurs viraux, des

phages, des plasmides, des phagemides, des cosmides, des phosmides, des vecteurs de type BAC, des bactériophages P1, des vecteurs de type BAC, des vecteurs de type YAC, des plasmides de levure ou encore tout autre vecteur connu de l'état de la technique par l'homme du métier.

On aura avantageusement recours selon l'invention à des vecteurs permettant une expression stable des acides nucléiques d'une banque d'ADN. A cette fin, de tels vecteurs incluent préférentiellement des séquences de régulation de la transcription qui sont localisées en phase ("operably linked") avec l'insert génomique de manière à permettre l'initiation et/ou la régulation de l'expression d'au moins une partie dudit insert d'ADN.

Il résulte de ce qui précède, que l'invention concerne encore un procédé de préparation d'une collection de vecteurs recombinants caractérisé en ce que les acides nucléiques obtenus à l'étape II-(iv) ou à l'étape I-(c) ou toute autre étape ultérieure d'un procédé de préparation d'une collection d'acides nucléiques à partir d'un échantillon de sol contenant des organismes selon l'invention sont insérés dans un vecteur de clonage et/ou d'expression.

Préalablement à leur insertion dans un vecteur de clonage et/ou d'expression, les acides nucléiques constitutifs d'une collection d'acides nucléiques selon l'invention peuvent être séparés en fonction de leur taille, par exemple par électrophorèse sur un gel d'agarose, le cas échéant après digestion à l'aide d'une endonucléase de restriction.

Selon un autre aspect, la taille moyenne des acides nucléiques constitutifs d'une collection d'acides nucléiques selon l'invention peut être rendue d'une taille sensiblement uniforme par la mise en oeuvre d'une étape de rupture physique préalablement à leur insertion dans le vecteur de clonage et/ou d'expression.

Une telle étape de rupture physique ou mécanique des acides nucléiques peut consister en des passages successifs de ces derniers, en solution, dans un canal métallique d'environ 0,4 mm de diamètre, par exemple le canal d'une aiguille de seringue ayant un tel diamètre.

La taille moyenne des acides nucléiques peut dans ce cas être comprise entre 30 et 40 kb de longueur.

La construction des vecteurs préférés selon l'invention est schématisée dans les figures 25 (cosmide intégratif conjugatif) et 26 (BAC intégratif).

Des vecteurs de clonage et/ou d'expression pouvant être
5 avantageusement utilisés aux fins d'insertion des acides nucléiques
contenus dans une collection ou banque d'ADN selon l'invention sont
notamment les vecteurs décrits dans le brevet européen N°EP-0 350
341 et dans le brevet US N°5 688 689, de tels vecteurs étant
spécialement adaptés à la transformation de souches d'actinomycètes.
10 De tels vecteurs contiennent, outre une séquence d'ADN de l'insert, une
séquence d'attachement att ainsi qu'une séquence d'ADN codant pour
une intégrase (séquence int) fonctionnelle dans les souches
d'actinomycètes.

Toutefois, il a été observé selon l'invention que certains
15 vecteurs de clonage et/ou d'expression présentaient des inconvénients
et que leur capacité fonctionnelle théorique n'était pas atteinte dans la
pratique.

Ainsi, il est apparu que le système d'intégration contenu dans
des vecteurs de l'état de la technique, et notamment dans les vecteurs
20 décrits dans le brevet européen n°EP 0 350 41 ne permettait pas en
réalité une bonne intégration de l'insert d'ADN de la banque au sein du
chromosome bactérien.

Partant de l'hypothèse que les déficits fonctionnels d'intégration
25 de tels vecteurs au sein du chromosome bactérien étaient dus à un
défaut dans l'expression du gène de l'intégrase présent dans ces
vecteurs, le demandeur a tout d'abord cherché à augmenter l'expression
du gène de l'intégrase en substituant au promoteur de la transcription
initial un promoteur de la transcription susceptible d'augmenter
30 significativement le nombre de transcrits de l'intégrase.

Les résultats ont été décevants et la fonction d'intégration au
chromosome de ces vecteurs n'a pas été améliorée.

De manière surprenante, il a été montré selon l'invention que
les difficultés d'expression de l'intégrase contenues dans cette famille de

vecteurs intégratifs ne se situait pas au niveau de la quantité d'expression des transcrits, mais au niveau de leur stabilité.

Selon une seconde hypothèse, le demandeur a pu montrer que le défaut de stabilité des transcrits de l'intégrase était causé par des déficits dans la terminaison de la transcription de l'ARN messenger correspondant.

Le demandeur a alors inséré un site terminateur placé en aval de la séquence codant pour l'intégrase du vecteur de manière à obtenir un ARN messager de taille déterminée. L'insertion d'un signal de terminaison additionnel en aval de la séquence nucléotidique codant pour l'intégrase du vecteur a permis l'obtention d'une famille de vecteurs intégratifs de type cosmide et de type BAC .

Préférentiellement, le site terminateur est placé en aval du site d'attachement att.

En outre, le demandeur a mis au point de nouveaux vecteurs conjugatifs et de nouveaux vecteurs réplicatifs du type cosmide et de nouveaux vecteurs conjugatifs de type BAC qui peuvent avantageusement être utilisés pour l'insertion des acides nucléiques constitutifs d'une collection d'acides nucléiques préparés selon le procédé de l'invention.

Lorsque l'insertion de fragments d'ADN de taille moyenne est recherchée, on utilise préférentiellement des vecteurs du type cosmide, capables de recevoir des inserts ayant une taille maximale d'environ 50 kb.

De tels vecteurs cosmiques sont tout particulièrement adaptés pour l'insertion d'acides nucléiques constitutifs d'une collection d'acides nucléiques obtenus selon le procédé de l'invention comprenant une première étape d'extraction directe d'ADN par lyse mécanique des organismes contenus dans l'échantillon de sol initial.

Lorsque l'insertion d'acides nucléiques de grande taille, en particulier d'acides nucléiques d'une taille supérieure à 100 kb, voire supérieure à 200, 300, 400, 500 ou 600 kb est recherchée, on aura alors recours préférentiellement à des vecteurs du type BAC capables de recevoir des inserts d'ADN d'une telle taille.

De tels vecteurs de type BAC sont tout particulièrement adaptés pour l'insertion des acides nucléiques constitutifs d'une collection d'acides nucléiques obtenus conformément au procédé selon l'invention dans lequel la première étape est constituée d'une extraction indirecte de l'ADN par séparation préalable des organismes contenus dans l'échantillon de sol initial et élimination des macro-constituants dudit échantillon de sol.

En particulier, des vecteurs du type BAC sont avantageusement mis en oeuvre pour l'insertion d'acides nucléiques de grande taille contenant, au moins partiellement, la séquence nucléotidique d'un opéron.

Ainsi, le procédé de préparation d'une collection de vecteurs recombinants de clonage et/ou d'expression selon l'invention est en outre caractérisé en ce que le vecteur de clonage et/ou d'expression est du type plasmide.

Selon un autre aspect, un tel procédé est caractérisé en ce que le vecteur de clonage et/ou d'expression est du type cosmide.

Selon un premier aspect, il peut s'agir d'un cosmide répliquatif chez *E.coli* et intégratif chez *Streptomyces*. Un vecteur cosmidique tout à fait préféré répondant à une telle définition est le cosmide pOS7001 décrit à l'exemple 3.

Selon encore un autre aspect, le vecteur cosmidique est conjuguatif et intégratif chez *Streptomyces*.

De manière générale, des vecteurs conjuguatifs de type cosmide ou de type BAC, qui comprennent dans leurs séquences nucléotidiques un motif reconnu par la machinerie enzymatique cellulaire appelé "origine de conjugaison" sont utilisés chaque fois que l'on veut éviter un recours à des techniques de transformation lourdes et peu automatisables.

Par exemple, la transfection de vecteurs initialement hébergés par des cellules de *E.coli* dans des cellules de *Streptomyces* nécessite classiquement une étape de récupération du vecteur recombinant contenu dans les cellules de *Escherichia coli*, et sa purification préalable à l'étape de transformation de protoplastes de *Streptomyces*. Il est communément admis qu'une transfection d'un ensemble de 1000 clones

de *Escherichia coli* dans *Streptomyces* requiert l'obtention d'environ 8000 clones pour que chaque clone de *E. coli* ait une chance d'être représenté.

5 A l'inverse, une étape de transfection par conjugaison d'un vecteur hébergé par *E.coli* vers des cellules de *Streptomyces* nécessite le même nombre de clones de chacun des micro-organismes, l'étape de conjugaison ayant lieu " clone à clone " et ne comprenant en outre pas les difficultés techniques liées à l'étape de transfert de matériel génétique par transformation de protoplastes, par exemple en présence
10 de polyéthylène glycol.

Afin d'optimiser la construction de banque d'ADN chez *Streptomyces*, il a été mis au point selon l'invention, de nouveaux vecteurs conjuguatifs de type cosmide et de type BAC de nature à permettre une efficacité maximale de l'étape de conjugaison.

15 Notamment, les nouveaux vecteurs conjuguatifs selon l'invention ont été construits en plaçant un gène marqueur de sélection à l'extrémité de l'ADN du vecteur qui est transféré à la bactérie réceptrice en dernier lieu. Ce perfectionnement aux vecteurs conjuguatifs de l'état de la technique permet de ne sélectionner positivement que les bactéries
20 réceptrices ayant reçu la totalité de l'ADN du vecteur et, en conséquence, la totalité de l'ADN de l'insert d'intérêt.

Des cosmides conjuguatifs et intégratifs chez *Streptomyces* préférés selon l'invention sont les cosmides pOSV303, pOSV306 et pOSV307 décrits à l'exemple 5.

25 Selon un autre aspect, un procédé de préparation d'une collection de vecteurs recombinants selon l'invention est mis en oeuvre à l'aide d'un cosmide réplcatif à la fois chez *E.coli* et chez *Streptomyces*. Un tel cosmide est avantageusement le cosmide pOS 700R décrit à l'exemple 6.

30 Selon encore un autre aspect, le procédé ci-dessus peut être mis en oeuvre avec un cosmide réplcatif chez *E. coli* et *Streptomyces* et conjuguatif chez *Streptomyces*.

Un tel cosmide réplcatif et conjuguatif peut être obtenu à partir d'un cosmide réplcatif conforme à l'invention, par l'insertion d'une

origine de transfert appropriée, telle que RK2, comme décrit à l'exemple 5 pour la construction du vecteur pOSV303.

Selon un autre mode de réalisation avantageux du procédé de préparation d'une collection de vecteurs recombinants selon l'invention, on a recours à un vecteur de clonage et/ou d'expression de type BAC.

Selon un premier aspect, le vecteur du type BAC est intégratif et conjugatif chez *Streptomyces*.

De manière tout à fait préférée, un tel vecteur BAC intégratif et conjugatif chez *Streptomyces* est le vecteur BAC pOSV 403 décrit à l'exemple 8, ou encore les vecteurs BAC pMBD-1, pMBD-2, pMBD-3, pMBD-4, pMBD-5 et pMBD-6 décrits à l'exemple 15.

L'invention a en outre pour objet un vecteur recombinant caractérisé en ce qu'il est choisi parmi les vecteurs recombinants suivants:

- a) un vecteur comprenant un acide nucléique constitutif d'une collection d'acides nucléiques selon l'invention;
- b) un vecteur tel qu'obtenu selon un procédé éliminant tout recours à l'action d'une endonucléase de restriction sur le fragment d'ADN à insérer, tel que décrit précédemment.

De manière tout à fait préférée, l'invention est également relative à un vecteur choisi parmi les vecteurs suivants:

- le cosmide pOS700I;
- le cosmide pOSV303;
- le cosmide pOSV306;
- le cosmide pOSV307;
- le cosmide pOS700R;
- le vecteur BAC pOSV403;
- le vecteur BAC pMBD-1;
- le vecteur BAC pMBD-2;
- le vecteur BAC pMBD-3;
- le vecteur BAC pMBD-4;
- le vecteur BAC pMBD-5;
- le vecteur BAC pMBD-6.

L'invention est en outre relative à une collection de vecteurs recombinants tels qu'obtenus selon l'un quelconque des procédés selon l'invention.

5 **Procédé de préparation d'un vecteur recombinant de clonage et/ou d'expression selon l'invention.**

Les techniques conventionnelles d'insertion d'ADN au sein d'un vecteur afin de préparer un vecteur de clonage et/ou d'expression recombinant font classiquement appel à une première étape au cours de laquelle une endonucléase de restriction est incubée à la fois avec l'ADN à insérer et avec le vecteur récepteur créant ainsi des extrémités compatibles entre l'ADN à insérer et l'ADN du vecteur permettant l'assemblage des deux ADN avant une étape de ligation finale permettant l'obtention du vecteur recombinant.

Toutefois, une telle technique conventionnelle présente des inconvénients notables, tout particulièrement lorsque est recherchée l'insertion d'acides nucléiques de grande taille dans un vecteur de clonage et/ou d'expression.

En effet, l'action préalable d'une enzyme de restriction sur les fragments d'ADN destinés à être insérés dans un vecteur est susceptible de réduire notablement la taille de cet ADN préalablement à son insertion dans le vecteur. Il va sans dire qu'une réduction significative de la taille de l'ADN préalablement à son insertion sur un vecteur est une situation particulièrement défavorable lorsqu'est recherché le clonage de fragments d'ADN de grande taille susceptible de contenir l'ensemble des séquences codantes et, le cas échéant, également des séquences régulatrices, d'un opéron dont l'expression constitue une voie de biosynthèse complète d'un métabolite d'intérêt industriel, et tout particulièrement d'un composé d'intérêt thérapeutique.

Pour remédier aux inconvénients des techniques de l'art antérieur, il a été mis au point selon l'invention deux procédés de préparation d'un vecteur recombinant de clonage et/ou d'expression qui ne nécessitent pas le recours à une endonucléase de restriction sur l'ADN à insérer préalablement à son introduction au sein du vecteur. De

tels procédés sont en conséquence tout à fait adaptés au clonage de longs fragments d'ADN susceptibles de contenir, au moins partiellement, l'ensemble des séquences codantes et, le cas échéant, également des séquences régulatrices, d'un opéron complet responsable d'une voie de biosynthèse.

Selon un premier aspect, un procédé de préparation d'un vecteur recombinant de clonage et/ou d'expression selon l'invention est caractérisé en ce que l'insertion d'un acide nucléique dans le vecteur de clonage et/ou d'expression, comprend les étapes suivantes:

- ouvrir le vecteur de clonage et/ou d'expression à un site de clonage choisi, à l'aide d'une endonucléase de restriction appropriée;

- ajouter un premier acide nucléique homopolymérique à l'extrémité 3' libre du vecteur ouvert;

- ajouter un second acide nucléique homopolymérique, de séquence complémentaire au premier acide nucléique homopolymérique, à l'extrémité 3' libre de l'acide nucléique à insérer dans le vecteur;

- assembler l'acide nucléique du vecteur et l'acide nucléique par hybridation du premier et du second acide nucléique homopolymérique de séquences complémentaires l'une de l'autre;

- refermer le vecteur par ligation.

Un tel procédé est décrit aux exemples 10 et 13 ci-après.

De manière avantageuse, le procédé ci-dessus peut comporter les caractéristiques suivantes, isolément ou en combinaison:

- le premier acide nucléique homopolymérique est de séquence poly(A) ou poly(T);

- le second acide nucléique homopolymérique est de séquence poly(T) ou poly(A).

- De manière tout à fait préférée, les acides nucléiques
- 5 homopolymériques ont une longueur comprise entre 25 et 100 bases nucléotidiques, préférentiellement entre 25 et 70 bases nucléotidiques.

Le procédé de préparation d'un vecteur recombinant de clonage et/ou d'expression décrit ci-dessus est particulièrement adapté à la construction de banques d'ADN dans des vecteurs de type BAC.

10 Ainsi, selon un mode de réalisation avantageux du procédé de préparation d'un vecteur recombinant décrit ci-dessus, ledit procédé est en outre caractérisé en ce que la taille de l'acide nucléique à insérer est d'au moins 100 kb, et préférentiellement d'au moins 200, 300, 400, 500 ou 600 kb.

- 15 Un tel procédé de préparation est donc particulièrement adapté à l'insertion des acides nucléiques contenus dans une collection d'acides nucléiques obtenus selon le procédé de l'invention.

Afin de permettre l'insertion de fragments d'ADN de grande taille dans des vecteurs de clonage et/ou d'expression, il a été mis au

20 point selon l'invention, un second procédé ayant permis d'éliminer tout recours à l'action d'une endonucléase de restriction sur l'ADN destiné à être inséré au sein du vecteur.

Un tel procédé de préparation d'un vecteur recombinant de clonage et/ou d'expression selon l'invention est caractérisé en ce que

25 l'étape d'insertion d'un acide nucléique dans ledit vecteur de clonage et/ou d'expression comprend les étapes suivantes:

- création de bouts francs sur les extrémités de l'acide nucléique de la collection par élimination des séquences 3' sortantes et
- 30 remplissage des séquences 5' sortantes;

- ouverture du vecteur de clonage et/ou d'expression à un site de clonage choisi à l'aide d'une endonucléase de restriction appropriée;

- 35 - addition d'adaptateurs oligonucléotidiques complémentaires ;

- création de bouts francs aux extrémités de l'acide nucléique du vecteur par élimination des séquences 3' sortantes et remplissage des séquences 5' sortantes, puis déphosphorylation des extrémités 5' afin de prévenir une recircularisation du vecteur;

- insertion de l'acide nucléique de la collection dans le vecteur par ligation.

De manière préférée, l'élimination des séquences 3' sortantes est réalisée à l'aide d'une exonucléase, telle que l'enzyme de Klenow.

De manière préférée, le remplissage des séquences 5' sortantes est réalisé à l'aide d'une polymérase, et de manière tout à fait préférée de la T4 polymérase, en présence des quatre nucléotides triphosphates.

Un procédé de préparation d'un vecteur recombinant de clonage et/ou d'expression par élimination des séquences 3' sortantes et remplissage des séquences 5' sortantes tel que décrit ci-dessus est particulièrement adapté à la construction de banques d'ADN à partir de vecteurs de type cosmide.

Un tel procédé d'obtention de vecteurs recombinants est décrit à l'exemple 12.

Dans un mode particulier de préparation d'un vecteur recombinant selon l'invention, des oligonucléotides comprenant un ou plusieurs sites de restriction rares sont ajoutés sur le vecteur au niveau du site de clonage de l'ADN à insérer, conformément à l'enseignement de l'exemple 10. Cet ajout d'oligonucléotides facilite la récupération ultérieure des inserts sans clivage de ces derniers.

CELLULES HOTES

Bien que tout type de cellules hôtes puisse être utilisé pour la transfection ou la transformation avec un acide nucléique ou un vecteur recombinant selon l'invention, notamment une cellule hôte procaryote ou eucaryote, on utilisera de préférence des cellules hôtes dont les

caractères physiologiques, biochimiques et génétiques sont bien caractérisés, facilement cultivables à grande échelle et dont les conditions de culture pour la production de métabolites soient bien connues.

5 De manière préférentielle, la cellule hôte réceptrice d'un acide nucléique ou d'un vecteur recombinant selon l'invention est phylogénétiquement proche des organismes donneurs contenus initialement dans l'échantillon de l'environnement desquels les acides nucléiques sont originaires.

10 De manière tout à fait préférée, une cellule hôte selon l'invention doit posséder un usage des codons similaire, ou du moins proche, des organismes donneurs présents initialement dans l'échantillon de l'environnement, tout particulièrement de l'échantillon de sol.

15 La taille des fragments d'ADN susceptible de porter les séquences nucléotidiques d'intérêt recherchées peut être variable. Ainsi, des enzymes codées par des gènes de taille moyenne de 1 kb pourront être exprimées à partir d'inserts de petite taille alors que l'expression de métabolites secondaires nécessiteront le maintien dans l'organisme hôte
20 de fragments de taille bien supérieure, par exemple de 40 kb à plus de 100 kb, 200 kb, 300 kb, 400 kb ou 600 kb.

Ainsi, les cellules hôtes de *Escherichia coli* constituent un choix privilégié pour le clonage de grands fragments d'ADN.

25 De manière tout à fait préférée, on aura recours à l'utilisation de la souche de *Escherichia coli* désignée DH10B et décrite par Shizuya et al; (1992) pour laquelle des protocoles de clonage dans des vecteurs BAC ont été optimisés.

Toutefois, d'autres souches de *Escherichia coli* peuvent être avantageusement utilisées pour la construction d'une banque d'ADN
30 selon l'invention, telles que les souches *E.coli* Sure, *E.coli* DH5 α , ou encore *E.coli* 294 (ATCC N°31446).

En outre, la construction d'une banque d'ADN par transfection de cellules de *E.coli* avec des vecteurs recombinants selon l'invention est également possible, l'expression de gènes de divers procaryotes tels

que *Bacillus*, *Thermotoga*, *Corynebacterium*, *Lactobacillus* ou *Clostridium* ayant été décrite dans la demande PCT N°WO 99/20799.

De manière générale, des cellules hôtes de *E.coli* peuvent dans
5 tous les cas constituer des hôtes transitoires dans lesquels des vecteurs
recombinants selon l'invention pourront être maintenus avec une grande
efficacité, le matériel génétique pouvant être facilement manipulé et
archivé et façon stable.

Dans le but d'exprimer la plus grande diversité moléculaire
10 possible, d'autres hôtes cellulaires pourront être également
avantageusement mis en oeuvre tels que des cellules de *Bacillus*,
Pseudomonas, *Streptomyces*, *Myxococcus*, *Aspergillus nidulans* ou
encore *Neurospora crassa*.

Il a en outre été montré selon la présente invention, que des
15 cellules de *Streptomyces lividans* peuvent être utilisées avec succès et
constituent des systèmes d'expression complémentaires à *Escherichia*
coli.

Streptomyces lividans constitue un modèle pour l'étude de la génétique
des *Streptomyces* et a également été utilisé comme hôte d'expression
20 hétérologue de nombreux métabolites secondaires. *Streptomyces*
lividans, possède en commun avec d'autres actinomycètes tels que
Streptomyces coelicolor, *Streptomyces griseus*, *Streptomyces fradiae*,
ainsi que *Streptomyces griseochromogenes*, les molécules précurseurs
et les systèmes de régulation nécessaires à l'expression de tout ou
25 partie des voies de biosynthèses complexes, telles que par exemple la
voie de biosynthèse des polykétides ou encore la voie de biosynthèse
des polypeptides non ribosomiques représentant des classes de
molécules de structures très diverses.

Streptomyces lividans présente également l'avantage
30 d'accepter l'ADN étranger avec des efficacités de transformation
élevées.

Ainsi, l'invention concerne aussi une cellule hôte recombinante
comprenant un acide nucléique selon l'invention, constitutif d'une
collection d'acides nucléiques préparée selon un procédé conforme à

l'invention, ou encore une cellule hôte recombinante comprenant un vecteur recombinant tel que défini précédemment.

Selon un premier aspect, il peut s'agir d'une cellule hôte recombinante d'origine procaryote ou eucaryote.

5 Avantageusement, une cellule recombinante selon l'invention est une bactérie, et de manière tout à fait préférée une bactérie choisie parmi *E.coli* et *Streptomyces*.

 Selon un autre aspect, une cellule hôte recombinante selon l'invention est caractérisée en ce qu'il s'agit d'une levure ou encore d'un
10 champignon filamenteux.

 L'invention a également trait à une collection de cellules hôtes recombinantes, chacune des cellules hôtes constitutive de la collection comprenant un acide nucléique provenant d'une collection d'acides nucléiques réalisée conformément à un procédé de préparation d'une
15 collection d'acides nucléiques à partir d'un échantillon de sol contenant des organismes tel que décrit ci-dessus.

 L'invention est également relative à une collection de cellules hôtes recombinantes, chacune des cellules hôtes constitutives de la collection comprenant un vecteur recombinant selon l'invention.

20 En raison de la grande taille des inserts il est nécessaire d'avoir une efficacité maximale de transformation. Dans ce but, une souche réceptrice de *Streptomyces lividans* exprimant l'intégrase de pSAM2 de façon constitutive afin de favoriser l'intégration site-spécifique du vecteur est préférée. Pour cela, le gène *int* sous contrôle d'un promoteur fort est
25 intégré dans le chromosome. La surproduction d'intégrase n'induit pas de phénomènes d'excision (Raynal et al., 1998).

 La production d'un nouveau métabolite à partir de l'insert pourrait être toxique pour *Streptomyces* si l'insert ne contient pas de gènes de résistance à l'antibiotique produit ou si ce gène est peu ou pas
30 exprimé. La capacité des différents gènes permettant à *Streptomyces ambifaciens* de résister à l'antibiotique qu'il produit est étudiée (Gourmelen et al., 1998; Pernodet et al., 1999). Certains de ces gènes codent des transporteurs de type ABC susceptibles de conférer un large spectre de résistance. Ces gènes peuvent être introduits et surexprimés
35 dans la souche hôte de *Streptomyces lividans*.

A l'inverse, une souche hypersensible aux antibiotiques peut être utilisée (Pernodet et al., 1996), afin de détecter dans la banque la présence de gènes de résistance. En effet, chez les micro-organismes producteurs d'antibiotique, ces gènes de résistance sont souvent associés aux gènes de la voie de biosynthèse de l'antibiotique. La sélection de clones résistants peut permettre d'effectuer simplement un premier tri avant les tests plus complexes de détection d'un nouveau métabolite produit par le clone.

10 **ISOLEMENT ET CARACTERISATION DE NOUVELLES SEQUENCES**
NUCLEOTIDIQUES CODANT POUR DES POLYKETIDES
SYNTHESES.

Selon l'invention, une collection de cellules hôtes recombinantes a été obtenue après transfection des cellules hôtes par une collection de vecteurs recombinants contenant chacun un insert d'acide nucléique provenant d'une collection d'acides nucléiques préparée conformément au procédé selon l'invention.

Plus précisément, les fragments d'ADN obtenus selon le procédé de l'invention dans lequel il est mis en oeuvre une étape d'extraction indirecte d'ADN des organismes contenus dans l'échantillon de sol ont été tout d'abord clonés dans le cosmide intégratif pOS700I.

L'étape d'insertion des fragments d'ADN dans le cosmide intégratif pOS700I a été réalisée selon le procédé de l'invention dans lequel des queues de polynucléotides homopolymériques poly(A) et poly(T) ont été ajoutées à l'extrémité 3' respectivement de l'acide nucléique du vecteur et des fragments d'ADN à insérer.

Les vecteurs recombinants ainsi construits ont été encapsidés dans des têtes de phage lambda et les phages obtenus ont été utilisés pour infecter des cellules de *E. coli* selon des techniques bien connues de l'homme du métier.

Une banque d'environ 5000 clones de *Escherichia coli* a été obtenue.

Cette banque de clones a été criblée avec des couples d'amorces spécifiques d'une séquence nucléotidique codant pour une

enzyme impliquée dans la voie de biosynthèse des polykétides, l'enzyme PKS de type I, aussi désignée β -kétolacyl synthase.

On rappelle ici que les polykétides constituent une classe chimique d'une grande diversité structurale comprenant un nombre important de molécules d'intérêt pharmaceutique tels que la tylosine, la monensine, la ivermectine, l'érythromycine, la doxorubicine ou encore le FK506.

Les polykétides sont synthétisés par condensation de molécules d'acétate sous l'action d'enzymes appelées polykétide synthases (PKSs). Il existe deux types de polykétide synthases. Les polykétide synthases de type II sont impliquées en général dans la synthèse des antibiotiques aromatiques polycycliques et catalysent la condensation d'unités acétate de façon itérative.

Les polykétide synthases de type I sont impliquées dans la synthèse des polykétides macrocycliques ou macrolides et constituent des enzymes modulaires multifonctionnelles.

Compte-tenu de leur intérêt thérapeutique, il existe un besoin dans l'état de la technique d'isoler et de caractériser de nouvelles polykétides synthases qui peuvent être utilisées pour la production de nouveaux composés pharmaceutiques, notamment de nouveaux composés pharmaceutiques à activité antibiotique.

Le criblage de la banque de clones recombinants décrite ci-dessus à l'aide d'amorces PCR amplifiant sélectivement des séquences nucléotidiques codant pour des polykétide synthases de type I a permis d'identifier des clones recombinants contenant des inserts d'ADN comprenant une séquence nucléotidique codant pour de nouvelles polykétide synthases. Les séquences nucléotidiques codant pour ces nouvelles polykétides synthases sont référencées comme les séquences SEQ ID N°33 à SEQ ID N°44 et SEQ ID N°115 à SEQ ID N°120.

Un autre objet de l'invention consiste en un acide nucléique codant pour une nouvelle polykétide synthase I, caractérisé en ce qu'il comprend l'une des séquences nucléotidiques SEQ ID N°34 à SEQ ID N°44 et SEQ ID N°115 à SEQ ID N°120.

De préférence, un tel acide nucléique se présente sous une forme isolée et/ou purifiée.

L'invention concerne aussi un vecteur recombinant comprenant un polynucléotide comprenant l'une des séquences SEQ ID N°34 à SEQ ID N°44 et SEQ ID N°115 à SEQ ID N°120

- L'invention a également trait à une cellule hôte recombinante
- 5 comprenant un acide nucléique choisi parmi les polynucléotides comprenant l'une des séquences nucléotidiques SEQ ID N°34 à SEQ ID N°44 et SEQ ID N° 115 à SDEQ ID N°120 ainsi qu'à une cellule hôte recombinante comprenant un vecteur recombinant dans lequel est
- 10 inséré un polynucléotide comprenant l'une des séquences nucléotidiques SEQ ID N°34 à SEQ ID N°44 et SEQ ID N°115 à SEQ ID N°120.

Avantageusement, les vecteurs recombinants contenant un insert d'ADN codant pour une nouvelle polykétide synthase de type I selon l'invention sont des vecteurs de clonage et d'expression.

- 15 De préférence, une cellule hôte recombinante telle que décrite ci-dessus est une bactérie, une levure ou encore un champignon filamenteux.

- Les séquences en acides aminés de nouvelles polykétide synthases provenant d'organismes contenus dans un échantillon de sol
- 20 ont été déduites des séquences nucléotidiques SEQ ID N°34 à SEQ ID N°44 ET SEQ ID N° 115 à SEQ ID N°120 ci-dessus. Il s'agit des polypeptides comprenant l'une des séquences en acides aminés SEQ ID N°48 à SEQ ID N°59 et SEQ ID N° 121 à 126.

- L'invention concerne encore de nouvelles polykétides synthases comprenant une séquence en acides aminés choisie parmi les séquences SEQ ID N°48 à SEQ ID N°59 et SEQ ID N° 121 à SEQ ID
- 25 N°126.

- Fait également partie de l'invention la séquence nucléotidique SEQ ID N°114 qui comprend six cadres ouverts de lecture qui codent
- 30 respectivement les polypeptides de séquences SEQ ID N°121 à SEQ ID N°126.

Fait également partie de l'invention la séquence nucléotidique SEQ ID N°113 du cosmide a26G1, qui contient la séquence complémentaire de la séquence SEQ ID N°114.

On a aussi extrait et amplifié selon l'invention de l'ADN génomique provenant de souches bactériennes pures, telles que *Streptomyces coelicolor* (ATCC N°101.478), *Streptomyces ambofaciens* (NRRL N°2.420), *Streptomyces lactamandurans* (ATCC N°27.382),
5 *Streptomyces rimosus* (ATCC N°109.610), *Bacillus subtilis* (ATCC N°6633) ou encore *Bacillus licheniformis* et *Saccharopolyspora erythraea*.

Une amplification par PCR de l'ADN de chacune des souches bactériennes décrites ci-dessus a été effectuée à l'aide des couples d'amorces spécifiques des séquences nucléiques de polykétide
10 synthase de type I.

De nouveaux gènes de polykétide synthases de type I bactériennes ont ainsi pu être isolés et caractérisés. Il s'agit des séquences nucléiques de séquences SEQ ID N°30 à SEQ ID N°32.

L'invention a donc en outre pour objet des séquences
15 nucléotidiques codant pour de nouvelles polykétides synthases de type I choisies parmi les polynucléotides comprenant l'une des séquences nucléotidiques SEQ ID N°30 à SEQ ID N°32.

Font également partie de l'invention des vecteurs recombinants comprenant les séquences nucléotidiques codant pour de nouvelles
20 polykétides synthases de type I définies ci-dessus.

L'invention concerne aussi des cellules hôtes recombinantes caractérisées en ce qu'elles contiennent un acide nucléique codant pour une nouvelle polykétide synthase de type I comprenant une séquence nucléotidique choisie parmi les séquences SEQ ID N°30 à SEQ ID N°32
25 ainsi que des cellules hôtes recombinantes comprenant un vecteur recombinant tel que défini ci-dessus.

L'invention a également pour objet des polypeptides codés par des séquences comprenant les acides nucléiques SEQ ID N° 30 à 32, et plus précisément des polypeptides comprenant les séquences d'acides
30 aminés SEQ ID N° 47 à SEQ ID N° 50.

L'invention a en outre pour objet un procédé de production d'une polykétide synthase de type I selon l'invention, ledit procédé de production comprenant les étapes suivantes:

- obtention d'une cellule hôte recombinante comprenant un acide nucléique codant pour une polykétide synthase de type I comprenant une séquence nucléotidique choisie parmi les séquences SEQ ID N°33 à SEQ ID N°44, SEQ ID N°30 à SEQ ID N°32 et SEQ ID N°115 à SEQ ID N°120;

- culture des cellules hôtes recombinantes dans un milieu de culture approprié;
- 10 - récupération et, le cas échéant, purification de la polykétide synthase de type I à partir du surnageant de culture ou du lysat cellulaire.

- Les nouvelles polykétide synthases de type I obtenues selon le
- 15 procédé décrit ci-dessus peuvent être caractérisées par fixation sur une colonne de chromatographie d'immuno-affinité sur laquelle des anticorps reconnaissant ces polykétides synthases ont été préalablement immobilisés.

- Les polykétide synthases de type I selon l'invention, et plus
- 20 particulièrement les polykétide synthases recombinantes décrites ci-dessus peuvent être aussi purifiées par des techniques de chromatographie liquide à haute performance (HPLC), telles que par exemple des techniques de chromatographie en phase inverse ou de chromatographie d'échanges d'anions ou de cations, bien connues de
- 25 l'homme du métier.

Les polykétide synthases, recombinantes ou non recombinantes, selon l'invention peuvent être utilisées pour la préparation d'anticorps.

- Selon un autre aspect, l'invention a donc encore pour objet un
- 30 anticorps reconnaissant spécifiquement une polykétide synthase de type I selon l'invention ou un fragment peptidique d'une telle polykétide synthase.

Les anticorps selon l'invention peuvent être monoclonaux ou polyclonaux. Les anticorps monoclonaux peuvent être préparés à partir

de cellules d'hybridome selon la technique décrite par KOHLER et MILSTEIN C. (1975), Nature, Vol.256:495.

Les anticorps polyclonaux peuvent être préparés par immunisation d'un mammifère, en particulier des souris, des rats ou des lapins avec une polykétide synthase de type I selon l'invention, le cas échéant en présence d'un composé adjuvant de l'immunité, tels que l'adjuvant complet de Freund, l'adjuvant incomplet de Freund, l'hydroxyde d'aluminium ou encore un composé de la famille des muramyl peptides.

Constituent également des " anticorps " au sens de la présente invention, les fragments d'anticorps tels que les fragments Fab, Fab', F(ab')₂, ou encore les fragments d'anticorps simple chaîne contenant la partie variable (ScFv) décrits par MARTINEAU et al. (1998) J. Mol. Biol., Vol.280 (1):117-127 ou encore dans le brevet US 4,946,778, ainsi que les anticorps humanisés décrits par REINMANN KA et al. (1997), AIDS Res. Hum. Retroviruses, vol.13(11):933-943 ou par LEGER O.J et al. (1997), Hum. Antibodies, vol.8 (1): 3-16.

Les préparations d'anticorps selon l'invention sont utiles notamment dans des tests immunologiques qualitatifs ou quantitatifs visant, soit à simplement détecter la présence d'une polykétide synthase de type I selon l'invention, soit à quantifier la quantité de cette polykétide synthase, par exemple dans le surnageant de culture ou le lysat cellulaire d'une souche bactérienne susceptible de produire une telle enzyme.

Un autre objet de l'invention consiste en un procédé de détection d'une polykétide synthase de type I selon l'invention ou un fragment peptidique de cette enzyme, dans un échantillon, ledit procédé comprenant les étapes de :

a) mettre en contact un anticorps selon l'invention avec l'échantillon à tester;

b) détecter le complexe antigène/anticorps éventuellement formé.

L'invention est également relative à un kit ou nécessaire de détection d'une polykétide synthase de type I selon l'invention dans un échantillon, comprenant :

a) un anticorps selon l'invention;

5 b) le cas échéant, des réactifs nécessaires à la détection du complexe antigène/anticorps éventuellement formé.

Un anticorps dirigé contre une polykétide synthase de type I selon l'invention peut être marqué à l'aide d'un marqueur détectable isotopique ou non isotopique, selon des procédés bien connus de
10 l'homme du métier.

Le criblage d'une banque d'ADN selon l'invention à l'aide d'une paire d'amorces hybridant avec des séquences cibles dont la présence est recherchée, telles que des séquences de la voie de biosynthèse de la puromycine, des séquences du gène *linA* impliquées dans la
15 biodégradation du lindane ou encore des séquences codant pour des polykétides synthases de type I ont été détaillées ci-avant.

L'invention a donc pour objet un procédé de détection d'un acide nucléique de séquence nucléotidique déterminée, ou de séquence nucléotidique structuralement apparentée à une séquence nucléotidique
20 déterminée, dans une collection de cellules hôtes recombinantes selon l'invention, caractérisé en ce qu'il comprend les étapes suivantes:

- mettre en contact la collection de cellules hôtes recombinantes avec un couple d'amorces hybridant avec la séquence nucléotidique déterminée ou hybridant avec la séquence nucléotidique
25 structurellement apparentée à une séquence nucléotidique déterminée;
- réaliser au moins trois cycles d'amplification ;
- détecter l'acide nucléique éventuellement amplifié.

Pour les conditions d'amplification appropriées en fonction des
30 séquences cibles recherchées, l'homme du métier pourra se référer avantageusement aux exemples ci-dessous.

Selon un autre aspect, l'invention concerne aussi un procédé de détection d'un acide nucléique, de séquences nucléotidiques déterminées, ou de séquences nucléotidiques structurellement
35 apparentées à une séquence nucléotidique déterminée, dans une

collection de cellules hôtes recombinantes selon l'invention, caractérisé en ce qu'il comprend les étapes suivantes:

- mettre en contact la collection de cellules hôtes recombinantes avec une sonde hybridant avec la séquence nucléotidique déterminée ou hybridant avec une séquence nucléotidique structurellement apparentée à la séquence nucléotidique déterminée;
- détecter l'hybride éventuellement formé entre la sonde et les acides nucléiques compris dans les vecteurs de la collection.

- 10 Pour effectuer le criblage d'une banque d'ADN selon l'invention en vue de détecter la présence d'une séquence nucléotidique codant pour un polypeptide capable de dégrader le lindane, on a détecté les clones recombinants d'intérêt par leur phénotype correspondant à leur capacité à dégrader le lindane. Dans ce but, les clones isolés et/ou des
- 15 ensembles de clones de la banque d'ADN préparée ont été mis en culture dans un milieu de culture en présence de lindane et la dégradation du lindane a été observée par la formation d'un halo trouble dans l'environnement immédiat des cellules.

- L'invention concerne aussi un procédé pour identifier la
- 20 production d'un composé d'intérêt par une ou plusieurs cellules hôtes recombinantes dans une collection de cellules hôtes recombinantes selon l'invention, caractérisé en ce qu'il comprend les étapes suivantes:

- culture des cellules hôtes recombinantes de la collection dans un milieu de culture approprié;
- détection du composé d'intérêt dans le surnageant de culture ou dans le lysat cellulaire d'une ou plusieurs des cellules recombinantes cultivées.

- L'invention a en outre pour objet un procédé pour sélectionner une cellule hôte recombinante produisant un composé d'intérêt dans une
- 30 collection de cellules hôtes recombinantes selon l'invention, caractérisé en ce qu'il comprend les étapes suivantes:

- culture des cellules hôtes recombinantes de la collection dans un milieu de culture approprié;

- détection du composé d'intérêt dans le surnageant de culture ou dans le lysat cellulaire d'une ou plusieurs des cellules hôtes recombinantes cultivées;

- sélection des cellules hôtes recombinantes produisant le
- 5 composé d'intérêt.

L'invention concerne encore un procédé pour la production d'un composé d'intérêt caractérisé en ce qu'il comprend les étapes suivantes:

- cultiver une cellule hôte recombinante sélectionnée selon le
- 10 procédé décrit ci-dessus;
- récupérer et, le cas échéant, purifier, le composé produit par ladite cellule hôte recombinante.

L'invention est également relative à un composé d'intérêt caractérisé en ce qu'il est obtenu selon le procédé ci-dessus décrit.

- 15 Un composé d'intérêt selon l'invention peut consister en un polykétide produit grâce à l'expression d'au moins une séquence nucléotidique comprenant une séquence choisie parmi les séquences SEQ ID N°33 à 44, SEQ ID N°30 à 32 et SEQ ID N°115 à SEQ ID N°120.

- 20 L'invention concerne encore une composition comprenant un polykétide produit grâce à l'expression d'au moins une séquence nucléotidique comprenant une séquence choisie parmi les séquences SEQ ID N°33 à SEQ ID N°44, SEQ ID N°30 à SEQ ID N°32, et SEQ ID N°115 à SEQ ID N°120.

- 25 Un polykétide produit grâce à l'expression d'au moins une séquence nucléotidique ci-dessus est préférentiellement le produit de l'activité de plusieurs séquences codantes incluses au sein d'un opéron fonctionnel dont les produits de traduction sont les différentes enzymes nécessaires à la synthèse d'un polykétide, l'une des séquences ci-
- 30 dessus étant comprise et exprimée dans ledit opéron. Un tel opéron comprenant une séquence d'acide nucléique selon l'invention codant pour une polykétide synthase peut être construit par exemple selon l'enseignement de Borchert et al. (1992).

- L'invention est encore relative à une composition
- 35 pharmaceutique comprenant une quantité pharmacologiquement active

d'un polykétide selon l'invention, le cas échéant en association avec un véhicule pharmaceutiquement compatible.

De telles compositions pharmaceutiques seront avantageusement adaptées pour l'administration, par exemple par voie parentérale, d'une quantité d'un polykétide synthétisé par une polykétide synthase de type I selon l'invention allant de $1\mu\text{g/kg}$ par jour à 10 mg/kg par jour, de préférence au moins $0,01\text{ mg/kg}$ par jour et de manière tout à fait préférée entre $0,01$ et 1 mg/kg par jour.

Les compositions pharmaceutiques selon l'invention peuvent être indifféremment administrées par voie orale, rectale, parentérale, intraveineuse, sous-cutanée ou encore intradermique.

L'invention concerne aussi l'utilisation d'un polykétide obtenu grâce à l'expression d'une polykétide synthase de type I selon l'invention pour la fabrication d'un médicament, en particulier d'un médicament à activité antibiotique.

L'invention sera en outre illustrée, sans pour autant être limitée, par les figures et les exemples ci-après.

La figure 1 illustre le schéma des différentes étapes de lyse effectuées selon les protocoles 1, 2, 3n 4a, 4b, 5a, et 5b décrits à l'exemple 1.

La Figure 2 illustre une électrophorèse sur gel d'agarose 0.8% des ADN extraits à partir de 300 mg du sol n°3 (Côte St André) après différents traitements de lyse (protocoles 1 à 5, cf. Fig. 1). M : marqueur de poids moléculaire de phage lambda

La Figure 3 illustre la proportion de différents genres d'actinomycètes cultivés à la suite des traitements 1 à 5 (cf. Fig. 1). Le nombre d'ufc (unité formant colonie) a été déterminé sur un milieu sélectif pour ce groupe de bactéries. Un nombre total d'environ 400 colonies a été analysé.

La Figure 4 illustre la récupération d'ADN de phage lambda digéré par *HindIII* additionné dans les sols à différentes concentrations avant (G) ou après (G*) broyage. Les traitements T (chocs thermiques)

et S (sonication) sont des traitements additionnels de lyse. La quantification a été réalisée par analyse au phospho-imageur après hybridation en dot-blot. Un échantillon de chaque sol a été utilisé pour chaque concentration de phage lambda ajouté. Les caractéristiques des sol sont reproduites dans le tableau 1. Les échantillons correspondant à 10 et 15 µg d'ADN ajouté n'ont pas été traités.

La Figure 5 illustre l'amplification par PCR des ADN extraits à partir de sol n°3 selon les protocoles 1, 2, 3, 5a et 5b. Les amorces FGPS 122 et FGPS 350 (tableau 2) ont été utilisés afin de cibler *Streptosporangium spp.* indigènes. Les extraits d'ADN ont été utilisés non dilués ou dilués au 1/10^{ème} et 1/100^{ème}. M : marqueur de poids moléculaires 123 pb (Gibco BRL), C : contrôle d'amplification sans ADN.

La Figure 6 illustre les quantités d'ADN extrait après inoculation de spores (a) ou de mycélium (b) de *S. lividans* OS48.3 inoculés dans les sols à différentes concentrations. La quantités de mycélium ajoutée dans le sol correspond au nombre de spores inoculées dans le milieu de germination. Environ 50% des spores ont germé, le nombre de cellules ou de génomes contenues dans les hyphes des spores germées n'a pas été déterminé. Les quantités de spores et de mycélium inoculées ne sont donc pas directement comparables. Le protocole d'extraction a été mené selon le protocole 6 (cf. section matériel et méthodes). Le symbole (') indique que de l'ARN a été inclus dans le tampon d'extraction. L'ADN cible a été amplifié par PCR avec les amorces FGPS 516 et FGPS 517, la quantification a été réalisée par phosphoimageur après hybridation en dot blot en utilisant le sonde FGPS 518. Un échantillon de chaque sol a été utilisé pour chaque concentration d'hyphes ou de spores. Les caractéristiques des sols sont décrites dans le tableau 1.

La figure 7 représente l'arbre phylogénétique obtenu par l'algorithme de Neighbour Joining , positionnant les séquences d'ADNr 16S contenues dans la banque d'ADN du sol, par rapport à des bactéries de références cultivées.

En grisé:..les séquences issues des pools de clones de la banque.

Les valeurs de bootstrap sont indiquées au niveau des noeuds, après rééchantillonnage de 100 répétitions. La barre d'échelle indique le nombre de substitutions par site. Le numéro d'accès des séquences dans la base de données Genbank est indiqué entre parenthèses.

La figure 8 représente un schéma du vecteur pOSint 1.

La figure 9 représente un schéma du vecteur pWED1.

La figure 10 représente un schéma du vecteur pWE15 (ATCC N° 37503).

La figure 11 représente un schéma du vecteur pOS 700I.

La figure 12 représente un schéma du vecteur pOSV010.

La figure 13 représente le fragment contenant un site "cos" inséré dans le plasmide pOSV010 au cours de la construction du vecteur pOSV 303.

La figure 14 représente un schéma du vecteur pOSV 303.

La figure 15 représente un schéma du vecteur pE116.

La figure 16 représente un schéma du vecteur pOS 700 R.

La figure 17 représente un schéma du vecteur pOSV 001.

La figure 18 représente le schéma du vecteur pOSV 002.

La figure 19 représente un schéma du vecteur pOSV 014.

La figure 20 représente un schéma du vecteur pBAC 11.

La figure 21 représente un schéma du vecteur pOSV 403.

La figure 22 représente les gels d'électrophorèse d'ADN de la banque après digestion par les enzymes BamHI et DraI des clones positifs de la banque criblée avec les oligonucléotides PKS-I.

La figure 23 illustre la production de puromycine par les recombinants de *S. lividans* comparée à la production de la souche sauvage *S. alboniger*.

10

La Figure 24 illustre l'alignement de PKSs du sol avec les sites actifs conservés d'autres PKSs. Les références pour chaque peptide sont indiquées. Les domaines bêta-kétoacyl synthase ont été alignés en utilisant le programme PILEUP de GCG (Wisconsin Package Version 9.1, Genetics Computer Group, Madison, Wisc).

15

La Figure 25 illustre la construction d'un cosmide intégratif conjugatif.

20

La Figure 26 illustre la construction d'un BAC intégratif conjugatif.

La figure 27 illustre le schéma de construction du vecteur pOSV 308.

25

La figure 28 illustre le schéma de construction du vecteur pOSV306.

La figure 29 illustre le schéma de construction du vecteur pOSV307.

30

La figure 30 illustre le schéma de construction du vecteur PMBD-1.

La figure 31 présente une carte détaillée du plasmide pMBD-2 ainsi qu'un schéma de construction du vecteur pMBD-3.

La figure 32 illustre une carte détaillée du plasmide pMBD-4.

La figure 33 illustre le schéma de construction du plasmide pMBD-5 à partir du plasmide pMBD-1.

La figure 34 illustre la carte détaillée du vecteur pBTP-3.

La figure 35 illustre le schéma de construction du vecteur pMBD-6 à partir du vecteur pMBD-1.

La figure 36 illustre la carte du cosmide a26G1 dont l'insert d'ADN contient des cadres ouverts de lecture codant pour plusieurs polykétides synthase.

La figure 37 est un schéma représentant l'insert d'ADN (brin +) du cosmide a26G1, sur lequel sont positionnés les différents cadres de lecture codant pour plusieurs polykétides synthase.

EXEMPLES:

EXEMPLE 1: Procédé de préparation d'une collection d'acides nucléiques à partir d'un échantillon de sol contenant des organismes, comprenant une étape d'extraction directe d'ADN à partir de l'échantillon de sol.

1. MATERIEL ET METHODES

1.1 SOLS: Les caractéristiques des six sols utilisés dans cette étude sont listées dans le tableau 1.

La teneur en argile et en matière organique va respectivement de 9 à 47% et de 1,7 à 4,7%, le pH variant de 4,3 à 5,8.

Des échantillons de sol ont été collectés à partir de la couche superficielle de 5 à 10 cm de profondeur. Toutes les racines visibles ont

été éliminées et les sols ont été conservés à 4°C pendant quelques jours si nécessaire, après quoi ils ont été séchés pendant 24 heures à la température ambiante et tamisés (taille moyenne de maille 2 mm) avant d'être conservés jusqu'à plusieurs mois à 4°C.

5

1.2 SOUCHES BACTERIENNES ET CONDITIONS DE CULTURE:

L'ADN extracellulaire ainsi que les souches bactériennes fournissant des cellules végétatives, des spores ou des hyphae, utilisées pour innoculer les échantillons de sol, ont été choisies de telle sorte que leur présence
10 puisse être suivie spécifiquement.

Afin d'obtenir de grandes quantités d'ADN extracellulaire, la souche lysogénique de *E.coli* 1192 Hfr P4X (metB), contenant le phage lambda CI857 Sam7, a été cultivée sur milieu Luria-Bertani (LB) pendant deux heures à 30°C, puis 30 minutes à 40°C, puis 3 heures à 37°C.
15 L'ADN du phage lambda a été extrait selon la technique décrite par SAMBROOK J. et al. (1989) Molecular Cloning: A Laboratory Manual, 2nd, ed. Cold Spring Harbor Laboratory, Cold Spring Harbor N.Y.

La souche avirulente de *Bacillus anthracis* (STERNE 7700) a été utilisée comme inoculum de cellules bactériennes. *Bacillus anthracis*
20 a été multiplié sur un bouillon de culture de type " trypticase soy broth " (TSB) (Biomérieux, Lyon, France) pendant environ 6 heures, en vérifiant que la DO₆₀₀ soit maintenue en dessous de 0,6. Ces conditions permettent le développement des cellules végétatives sans formation de spores (Patra et al., (1996), FEMS Immunol. Medical Microbiology, vol.15:223-231.). Les spores de *Streptomyces lividans* OS48.3 (CLERC-
25 BARDIN et al. non publié) ont été éliminées mécaniquement des cultures de l'organisme sur un milieu R2YE (HOPWOOD et al., (1985), Genetic Manipulation of Streptomyces-A Laboratory Manual. The John Innes Foundation , Norwich ,United Kingdom). Les hyphae de *S.lividans* OS48.3 ont été obtenus à partir des spores en pré-germination, car l'on
30 s'attendait à ce que l'utilisation de hyphae courtes minimise la rupture et la perte subséquente d'ADN. Les spores ont été mises en suspension dans du tampon TES⁻ (Acide N-Tris [hydroxyméthyl]méthyl-2-aminoéthanesulfonique ; Sigma-Aldrich Chimie, France) (0,05M; pH 8)
35 (Holben WE et al., (1988), APPL. Environ. Microbiol. vol.54:703-711,

puis ont été soumises à un choc thermique (50°C pendant 10 minutes suivi d'un refroidissement sous un courant d'eau froide puis ajoutées à un volume égal de milieu de pré-germination (extrait de levure 1%, casaminoacides 1% CaCl₂ 0,01 M).

- 5 La solution a été incubée à 37°C sur un agitateur. La proportion de spores germées a été estimée à environ 50%, en accord avec les résultats de HOPWOOD et al. (1985). Après centrifugation, les culots ont été resuspendus dans du tampon TES, ajoutés à 3% de milieu TSB, et incubés à 37°C jusqu'à l'obtention d'une DO₄₅₀ de 0,15 (HOPWOOD et al. (1985)). *Streptomyces hygroscopicus* SWN 736 et
- 10 *Streptosporangium fragile* AC1296 (Institute Pushino, Moscou) ont été cultivés selon des techniques décrites par HICKEY et TRESNER (1952).

- L'ADN des spores et des hyphae de *S. lividans* a été extrait à partir des cultures pures selon le protocole de lyse 6 décrit ci-dessous
- 15 (excepté qu'aucun broyage n'a été réalisé), tandis que les spores de *S. hygroscopicus* et de *S. fragile* ont été extraites par lyse chimique/enzymatique (Hintermann et al., 1981).

- 1.3 CHOIX DU TAMPON D'EXTRACTION:** Un tampon TENP (50 mM
- 20 Tris, 20 mM EDTA, 100 mM NaCl, 1% pds/vol de polyvinylpyrrolidone développé par PICARD (1992) a été utilisé. Des tampons similaires ont été ultérieurement utilisés par d'autres auteurs (CLEGG et al., 1997; KUSKE et al., 1998; ZHOU et al., 1996).

- Le Tris et l'EDTA protègent l'ADN de l'activité nucléase, le NaCl
- 25 apporte un effet dispersant et la PVPP absorbe les acides humiques et les autres composés phénoliques (HOLBEN et al. (1988); PICARD et al., (1992).

- Dans cette étude, l'efficacité d'extraction de ce tampon a été évaluée à différents pH (6,0 - 10,0) en utilisant 20 sols différents ayant
- 30 une gamme de pH de 5,8 à 8,3 et une teneur en matière organique entre 0,2 et 6,3%. Ces vingt sols (les autres caractéristiques ne sont pas indiquées) ont été utilisés uniquement dans cette expérience. La quantité d'ADN a été déterminée de manière colorimétrique comme décrit par RICHARD (1974), et détaillé ci-après.

1.4 PROTOCOLE DE LYSE *IN SITU* ET D'EXTRACTION D'ADN:

Plusieurs protocoles utilisant un nombre croissant d'étapes ont été testés afin d'évaluer l'efficacité de différentes techniques pour lyser les microbes du sol *in situ*. Pour ces expériences, la microflore indigène du sol a été ciblée dans six sols. Des expériences additionnelles ont été conduites afin d'étudier les effets des traitements de lyse sur l'ADN libéré, en analysant les quantités et la qualité d'ADN récupéré provenant d'un ADN de phage lambda préalablement additionné aux sols.

Une fois qu'un protocole optimisé (désigné protocole 6) a été développé, ce protocole a été utilisé pour quantifier l'ADN provenant d'*Actinomycètes* indigènes et d'ADN provenant de bactéries Gram-positives inoculées dans les sols sélectionnés. Dans tous les cas, les échantillons de sol ont été séchés et passés au tamis comme décrit ci-dessus.

Après broyage, 0,5 ml de tampon TENP ont été ajoutés à 200 mg poids sec de sol excepté pour le protocole 1 dans lequel le tampon a été ajouté à un sol non broyé).

Pour les divers traitements de lyse (voir ci-dessous), les suspensions de sol ont été passées au Vortex pendant dix minutes et centrifugées (4000 g pendant cinq minutes), après quoi une fraction aliquote (25 µl) du surnageant a été analysée par électrophorèse sur gel (0,8% d'agarose).

Une autre fraction aliquote du surnageant représentant un volume connu, généralement 350 µl, a été précipitée avec de l'isopropanol.

Cinq fractions aliquotes (représentant de l'ADN dérivé de 1 g de sol) ont été réunies et resuspendues dans 100 µl d'un tampon TE stérile (10 mM Tris, 1 mM EDTA, pH 8,0) avant purification (protocole D, voir ci-dessous) et quantification, soit par hybridation (Dot Blot) de l'ADN total, soit par hybridation (Dot Blot) des produits d'amplification PCR (voir ci-dessous).

Les signaux d'hybridation ont été quantifiés par imagerie par phosphorescence (technique de "phospho-imaging" voir ci-dessous).

1.5 EVALUATION DES METHODES DE LYSÉ CELLULAIRE *IN SITU*:

La qualité et la quantité de l'ADN extrait après un nombre croissant d'étapes de traitement de lyse (protocole 2-5b) ont été comparées à celles de l'ADN extracellulaire obtenu après lavage du sol avec un tampon d'extraction (protocole 1; voir aussi figure 1).

Protocole 1: Pas de traitement de lyse.

Le tampon TENP a été ajouté à un sol non broyé, une étape d'extraction d'ADN a été réalisée comme décrit ci-dessus.

Protocole 2. Broyage du sol suivi d'une extraction d'ADN.

Deux types de dispositifs différents ont été utilisés pour le broyage du sol.

Afin de comparer leur efficacité respective, 5g de sol sec ont été broyés pendant 30 secondes dans un broyeur contenant des anneaux de tungstène, ou pendant des temps variés jusqu'à 60 minutes dans un broyeur de sol contenant un mortier et des billes en agate (20 mm de diamètre).

Le tampon TENP est ensuite ajouté et l'ADN est extrait comme décrit ci-dessus.

Les résultats d'électrophorèse sur gel ont montré qu'un broyage de 40 minutes en utilisant des billes en agate étaient nécessaires afin d'obtenir des quantités d'ADN extraits équivalentes à celles obtenues après 30 secondes de broyage en utilisant des anneaux de tungstène.

La distribution de taille des fragments d'ADN est similaire quelle que soit la méthode employée.

Ainsi, ces traitements ont été considérés comme équivalents et celui qui sera utilisé dans les protocoles décrits ci-dessous ne sera en conséquence pas spécifié.

Dans les protocoles 3 à 5, l'efficacité de plusieurs autres traitements de lyse ultérieure au broyage du sol a été testée, soit séparément, soit dans différentes combinaisons.

Protocole 3:

- Ce protocole est identique au protocole 2 , sauf qu'il comprend une étape d'homogénéisation à l'aide d'un mixeur de type Ultraturax (Janker et Kunkel, IKA Labortechnik, Allemagne) réglé à la moitié de la vitesse maximale pendant 5 minutes.

PROTOCOLES 4a et 4b:

- Ces protocoles sont identiques au protocole 3 à l'exception d'une étape additionnelle de sonication.

- Deux types de dispositifs sonicateurs ont été comparés : un sonicateur à micropointe de titane (600W Vibracell Ultrasonicator, Bioblock, Illkirch, France) (Protocole 4a) et un sonicateur de type Cup Horn (protocole 4b).

La micropointe Vibracell produisant des ultrasons est en contact direct avec la solution de sol.

- En ce qui concerne le dispositif de type Cup Horn, la solution de sol est conservée dans des tubes qui sont placés dans un bain d'eau à travers lequel passent les ultrasons.

Des expériences préliminaires ont été réalisées afin de déterminer les conditions optimales pour les deux sonicateurs (résultats non présentés).

- Le meilleur compromis, en terme de quantité d'ADN extrait et de taille de fragments, consiste en une sonication avec la micropointe de titane et le sonicateur de type Cup Horn respectivement pendant 7 et 10 minutes, en réglant la puissance à 15 W et avec des cycles actifs à 50%.

Protocoles 5a et 5b:

30

Après sonication avec une micropointe de titane ou un dispositif de type Cup Horn (respectivement protocoles 4a et 4b) du lysozyme et de l'achromopeptidase ont été ajoutés, chacune des enzymes à une concentration finale de 0,3 mg/ml.

Les suspensions de sol ont été incubées pendant 30 minutes à 37°C, après quoi du lauryl sulfate à une concentration finale de 1 % a été ajouté, puis des suspensions ont été incubées pendant 1 heure à 60°C avant centrifugation et précipitation comme décrit ci-dessus.

5 En plus des protocoles décrits ci-dessus, l'effet de la sonication (Cup Horn, voir protocole 4b) et de chocs thermiques (30 secondes dans l'azote liquide suivi de trois minutes dans l'eau bouillante, les traitements étant répétés trois fois) sur l'ADN de phage lambda digéré par HindIII préalablement ajouté au sol ont été examinés (voir ci-après).

10 Des chocs thermiques ont été suggérés dans l'état de la technique comme des moyens de lyse cellulaire *in situ* (PICARD et al. (1992)). Cependant, du fait qu'un tel traitement a un effet préjudiciable sur l'ADN libre (voir la section résultats) il n'a pas été inclus dans les protocoles décrits ci-dessus.

15

PROTOCOLE OPTIMISE

Après évaluation des différents traitements de lyse, un protocole optimisé a été défini, désigné protocole 6. Le protocole 6 est
20 identique au protocole 5b excepté que, avant la sonication, les suspensions de sol sont soumises à un traitement par Vortex puis agitées par rotation sur une roue pendant deux heures avant d'être congelées à - 20°C.

Après décongélation, les suspensions de sol sont passées au
25 Vortex pendant 10 minutes avant sonication. Le protocole 6 a été utilisé dans les expériences dans lesquelles les sols ont été ensemencés avec des cellules bactériennes ainsi que dans les expériences dans lesquelles les actinomycètes indigènes ont été quantifiés (voir ci-dessous).

30 **1.6 COMPTAGE AU MICROSCOPE:** L'efficacité du broyage du sol comme méthode pour lyser des cellules bactériennes a été examinée au microscope.

5g de sol brut séché ont été mélangés dans un dispositif de type Waring Blender avec 50 ml d'eau stérilisée ultrapure pendant 1,5
35 minutes; simultanément, 1g (poids sec) de sol broyé (protocole n°2) a

été mis en suspension dans 10 ml par agitation pendant 10 minutes. Les suspensions de sol ont fait l'objet de dilutions en séries et de l'acridine orange a été ajoutée à une concentration finale de 0,001%.

Après 2 minutes, les suspensions ont été filtrées à travers une
5 membrane de marque NUCLEOPORE de type 0,2 µm black. Chaque filtre a été rincé avec de l'eau stérile lysée, traitée avec 1 ml d'isopropanol pendant 1 minute afin de fixer les cellules bactériennes, puis rincé de nouveau.

Les cellules bactériennes ont été comptées à l'aide d'un
10 microscope à épifluorescence du type Zeiss Universal avec un objectif 100x. Pour chacun des types de sol, trois filtres ont été comptés, et au moins 200 cellules ont été comptées sur chacun des filtres.

1.7 NUMERATION DES ACTINOMYCETES CULTIVABLES ET
15 **NOMBRE TOTAL D'UNITES FORMANT COLONIES (CFU):** Les actinomycètes ayant survécu aux traitements de lyse (protocoles 1-5) ont été examinés spécifiquement avec le sol n°3 (Côte Saint André, voir tableau 1).

Après une dilution de 10 fois d'une solution d'extrait de levure
20 (6% poids/volume) et de SDS (0,05%) afin d'induire la germination (Hayakawa et al. (1988)), les suspensions de sol ont été diluées en séries dans de l'eau stérile, incubées à 40°C pendant 20 minutes et ensemencées sur du milieu HV (HAYAKAWA et al., 1987).

Le milieu HV a été additionné de actidione (50 mg/l) et de
25 nystatine (50 mg/ml).

Les colonies d'actinomycètes ont été comptées après incubation pendant 15 jours à 28°C.

Au total, environ 400 colonies ont été examinées. L'identification a été réalisée sur la base des caractéristiques
30 morphologiques macro-et microscopiques ainsi que sur l'analyse de la teneur en acide diaminopimélique des isolats (SHIRLING et al., 1966); STANECK et al., 1974; WILLIAMS et al., 1993).

La quantité totale de bactéries cultivables (CFU totales) a été également déterminée pour chacun des protocoles de lyse 1 à 5. Les
35 suspensions de sol ont été diluées en série et ensemencées en triple sur

un milieu agar Bennett (WAKSMAN et al., 1961) additionné de nystatine et d'actidione (chacune à 50 mg/l).

- Chaque boîte de Pétri a été couverte d'un filtre de nitrate de cellulose (Millipore) et incubée pendant trois jours à 28°C. Après la
5 numération des colonies sur les membranes, les filtres ont été retirées et les boîtes de Pétri ont été à nouveau incubées pendant 7 jours à 28°C puis comptées à nouveau.

- 1.8 RECUPERATION DE L'ADN DE PHAGE LAMBDA AJOUTE AUX**
10 **SOLS:** L'ADN de phage lambda a été digéré avec HindIII, extrait par un mélange de phénol-chloroforme, précipité puis resuspendu dans de l'eau stérile ultrapure selon des protocoles standard (SAMBROOK et al., 1989).

- Des dilutions correspondant respectivement à 0, 2,5, 5, 7,5, 10
15 et 15 µg d'ADN/g de poids sec de sol ont été préparées dans des volumes de 60 µl. Ces dilutions d'ADN ont été ajoutées à des lots de 5g de sol sec qui ont été subséquemment vigoureusement mélangés par vortex pendant 5 minutes avant broyage.

- L'ADN de phage lambda a aussi été ajouté à un sol avant
20 broyage à des concentrations correspondant à 0, 10 et 15 µg d'ADN/g de poids sec du sol.

Après broyage, le tampon d'extraction est ajouté et l'ADN est extrait selon le protocole 2 (voir ci-dessus).

- 25 **1.9 SATURATION DES SITES D'ADSORPTION AVEC DE L'ARN:** Afin de déterminer si la saturation des sites d'adsorption d'acides nucléiques des colloïdes du sol pouvait augmenter le taux de récupération de l'ADN, le terreau sablonneux (sol n°4) et le sol argileux (sol n°5) ont été incubés avec une solution d'ARN avant tout autre traitement.

- 30 De l'ARN commercial de *Saccharomyces cerevisiae* (BOHRINGER MANNHEIM, MEYLAN, France) a été dilué dans du tampon phosphate (pH 7,1) et ajouté aux échantillons de sol sec et tamisés (2 ml/g de sol) à des concentrations finales de 20, 50 et 100 mg d'ARN/g de poids sec du sol.

Les tubes contenant les suspensions de sol ont été agités par rotation pendant deux heures à température ambiante. Après centrifugation, les culots de sol ont été séchés au four (50°C) pendant la nuit. L'ADN de phage lambda a ensuite été ajouté aux sols (0, 20 ou 50 µg/g de poids sec du sol) afin de simuler le sort de l'ADN libéré après lyse cellulaire.

L'ADN a été extrait selon le protocole n°2. Il a été déterminé par la suite qu'un effet identique de l'addition d'ARN sur la récupération d'ADN pouvait être atteint en ajoutant l'ARN directement au tampon d'extraction.

Cette procédure simplifiée a été utilisée pour le sol argileux n°5 dans les expériences dans lesquelles les micro-organismes ont été inoculés dans les sols.

L'ARN a ensuite été ajouté à une concentration correspondant à 50 mg d'ARN/g de poids sec du sol.

1.10 DETERMINATION QUALITATIVE ET QUANTITATIVE DE L'EFFICACITE DES PROTOCOLES D'EXTRACTION: La qualité de l'ADN (absence de dégradation) a été estimée sur la base de la taille des fragments d'ADN ou de la position relative des bandes de migration d'ADN après électrophorèse d'une fraction aliquote d'une solution d'ADN sur un gel d'agarose à 0,8%.

L'intensité de fluorescence a permis une estimation semi-quantitative des rendements d'extraction.

Une autre fraction aliquote a été utilisée pour des déterminations quantitatives de la teneur en ADN par hybridation (Dot Blot) et analyse au phospho-Imager. Le protocole d'hybridation sur tache a été décrit par SIMONET et al. (1990).

Les membranes d'hybridation (GeneScreen plus, Life Science Products, Boston, Etats-Unis d'Amérique) ont été préhybridées pendant au moins 2 heures dans 20 ml d'une solution contenant 6 ml de 20 x SSC, 1 ml de solution de DENHARDT's, 1 ml de SDS à 10% et 5 mg d'ADN de sperme de saumon.

L'hybridation a été réalisée pendant une nuit dans la même solution en présence d'une sonde marquée préalablement à deux

lavages des membranes dans un tampon SSC 2 x pendant 5 minutes à température ambiante, puis un troisième lavage dans du tampon SSC 2 x, SDS 0,1% et un quatrième lavage dans du tampon SSC 1 x, SDS 0,1% pendant 30 minutes à la température d'hybridation.

5 Les signaux d'hybridation ont été quantifiés avec un système d'imagerie radioanalytique BIORAD (Molecular Analyst Software, BIORAD, Ivry S/Seine, France).

Afin de quantifier la quantité totale d'ADN dérivée de la microflore indigène, les différents sols ont été extraits selon les
10 protocoles n°1 à 5. L'ADN non amplifié a été appliqué sur les membranes de Dot Blot et hybridé en utilisant la sonde universelle FGPS431 (tableau 2).

Cette sonde, qui hybride aux positions 1392-1406 du gène de l'ADNr 16S de *E.coli* (Amann et al. (1995)) a été marquée à ses
15 extrémités avec un ATP α -³²P en utilisant une polynucléotide kinase T4(BOEHRINGER MANNHEIM, Melan, France).

Une courbe de calibration a été préparée à partir de l'ADN de *E.coli* DH5 α . La conversion des calculs aux bactéries du sol a nécessité une simplification, partant de l'hypothèse que le nombre de copies
20 moyen (rrn) est de 7, comme pour *E.coli*.

L'ADN de phage lambda digéré par HindIII a été utilisé pour quantifier la récupération de l'ADN extracellulaire. Des extraits non amplifiés à partir de sols, auxquels de l'ADN de phage lambda avait été
ajouté, ont été hybridés avec de l'ADN de phage lambda digéré par
25 HindIII marqué au hasard en utilisant le fragment Klenow (Boehringer Mannheim, Melan, France).

Les quantités d'ADN ont été calculées par interpolation à partir d'une courbe de calibration préparée avec l'ADN purifié.

La quantité totale d'ADN extrait à partir des sols n°1, 2, 3, 4 et 6
30 selon le protocole n°2 (broyage) a également été quantifiée de manière colorimétrique selon la technique décrite par RICHARD (1974).

Brièvement, de l'ADN a été mélangé avec du HClO₄ concentré (la concentration finale de HClO₄ était de 1,5 N). On a mélangé 2,5 volumes de cette solution avec 1,5 volumes de DPA (diphénylamine,
35 Sigma-Aldrich, France) et laissé incuber le mélange à la température

ambiante pendant 18 heures, préalablement à la détermination de la DO à 600 nm. Les extraits d'ADN du sol ont été quantifiés par rapport à une courbe standard réalisée par l'ADN extrait à partir de *E.coli* DH5 α selon les protocoles standards (SAMBROOK et al., (1989)).

5

1.11 DEVELOPPEMENT D'UNE TECHNIQUE DE QUANTIFICATION D'ADN EN UTILISANT L'AMPLIFICATION PCR ET L'HYBRIDATION:

Pour les amplifications par PCR, de l'ADN polymérase Taq (Appligene Oncor, France) a été utilisé selon les instructions du fabricant.

- 10 Le programme PCR utilisé pour toutes les amplifications est le suivant: dénaturation initiale pendant 3 minutes à 95°C, puis 35 cycles consistant en 1 minute à 95°C, 1 minute à 55°C et 1 minute à 72°C, suivie par une extension finale à 72°C pendant 3 minutes.

- 15 L'ADN isolé et purifié à partir de *Streptosporangium fragile* a été utilisé comme témoin à des concentrations allant de 100 fg à 100 ng.

- Afin d'amplifier spécifiquement l'ADN de ce genre bactérien, on a choisi les amorces FGPS122 et FGPS350 (tableau 2), complémentaires à une partie de l'ADNr 16S, après alignement des séquences d'ADNr 16S d'actinomycètes. Leur spécificité a été testée
20 sur une collection de souches d'actinomycètes (*Streptomyces*, *Streptosporangium* et d'autres genres fortement apparentés).

- Les produits de PCR ont été hybridés avec la sonde oligonucléotidique FGPS643 (tableau 2). Afin de simuler le niveau de pureté obtenu en routine avec de l'ADN extrait à partir du sol, des
25 témoins d'ADN pur de *S. fragile* ont été mélangés avec les extraits de sol obtenus après des traitements selon les protocoles de lyse 4b et 5b puis purifiés selon le protocole D.

- Avant utilisation, les extraits de sol ont été traités avec de la DNase (une unité de DNase/ml, GIBCO BRL) pendant 30 minutes à
30 température ambiante. La DNase a ensuite été inactivée par chauffage à 65°C pendant 10 minutes. Une vérification de l'inactivation a été réalisée par PCR. Les concentrations d'acides humiques ont été mesurées par spectrophotométrie (DO_{280nm}) contre une courbe standard d'acides humiques commerciaux (Sigma).

Des solutions de sol traitées à la Dnase non diluées, diluées 10x et diluées x 100 ont été mélangées de 100. fg à 100ng d'ADN de *S. fragile* avant l'amplification par PCR. Dans une autre série d'expériences, les concentrations croissantes d'ADN de *Streptomyces hygroscopicus* de (100 pg à 1 µg) ont été ajoutées à l'ADN de *S. fragile* afin de simuler la présence d'ADN non-cible et son influence sur le procédé PCR.

1.12 PURIFICATION DES EXTRAITS D'ADN BRUT: Quatre méthodes de purification d'ADN ont été comparées. L'ADN a été extrait à partir de 1g (poids sec de sol selon le protocole 4a et remis en suspension dans 100 µl de tampon TE8 (50 mM Tris, 20 mM (EDTA, pH 8,0).

Protocole A

Elution à travers deux colonnes successives Elutip d (SCHLEICHER et SCHUELL, Dassel, Allemagne) (PICARD et al., (1992)).

Protocole B:

Elution à travers une colonne SEPHACRYL S200 (Pharmacia Biotech, Uppsala, Suède) suivie d'une élution à travers une colonne Elutip d (NESME et al. (1995)).

Protocole C:

Séparation à l'aide d'un système aqueux à deux phases avec 17,9% (poids/poids) de PEG 8000 (Merck, Darmstadt, Allemagne) et 14,3% (poids/poids) de $(\text{NH}_4)_2\text{SO}_4$ (ZASLAVSKY, (1995)).

Après un mélange vigoureux au vortex, les deux phases ont été laissées à température ambiante pour leur séparation.

1 ml de chacune des phases a été transféré dans un autre tube, mélangé avec 100µl de l'échantillon et laissé à 4°C pendant une nuit pour permettre la séparation.

La phase inférieure a été dialysée pendant une heure à travers une membrane Millipore en présence d'un excès d'un tampon TE 7,5 (10 mM Tris, 1 mM EDTA à pH 7,5 et 1 M Mg Cl₂) afin d'éliminer les sels en excès.

5

Protocole D:

Elution à travers une colonne de type Microspin Sephacryl S400 HR (Pharmacia Biotech, Uppsala, Suède) , suivie d'une élution à travers une colonne de type Elutip d.

Chaque protocole est terminé par une étape de précipitation à l'éthanol, et l'ADN est remis en suspension dans 10 µl de tampon TE 7,5. L'efficacité des protocoles de purification a été vérifiée par amplification PCR de fractions aliquotes non diluées des solutions d'ADN et de fractions aliquotes diluées 10 et x 100 fois, en utilisant des protocoles standard (voir ci-dessous).

1.13 RECUPERATION DE L'ADN A PARTIR DE MICROORGANISMES INNOCULES:

Les cellules, spores et hyphae ont été lavées deux fois et dénombrées par comptage sur plaque ou comptage microscopique direct. Des lots de 5g de sol sec et tamisé (sols n°2, 3 et 5) ont été inoculés avec 100 µl d'une suspension de spores et d'hyphae de *S. lividans* à des concentrations correspondant à 0,10³, 10⁵, 10⁷ et 10⁹ spores/g de poids sec de sol, ou avec des cellules végétatives de *B.anthraxis* à des concentrations correspondant à 0,10⁷ et 10⁹ cellules par gramme de poids sec du sol.

Les quantités de hyphae de *S. lividans* ont été calculées sur la base du nombre de spores desquelles elles sont originaires. Après addition des suspensions bactériennes, les échantillons de sol sont mélangés vigoureusement par vortex pendant 5 minutes avant broyage. L'ADN est extrait selon le protocole n°6 (voir ci-dessous).

L'amplification PCR suivie d'une hybridation sur tache (Dot Blot) et imagerie par phosphorescence (phospho-imaging) a été utilisée afin

de quantifier les quantités d'ADN récupérées à partir des cellules, des spores et du mycélium bactérien inoculé dans les sols.

L'extraction d'ADN a été réalisée selon le protocole de lyse n°6. L'amplification PCR et l'hybridation ont été réalisées comme décrit ci-dessus. Les amorces et les sondes sont ciblées sur des régions chromosomiques localisées en dehors de la région 16S, et sont hautement spécifiques des organismes respectifs, de manière à éviter des signaux de bruit de fond.

Pour les solsensemencés avec *B. anthracis*, les amorces R499 et R500 ont été utilisées (Patra et al. (1996)) et les produits d'amplification ont été hybridés avec la sonde oligonucléotidique C501 (tableau 2).

Pour les solsensemencés avec *S. lividans*, les réactions PCR ont été réalisées en utilisant les amorces FGPS516 et FGPS517, et les produits d'amplification ont été hybridés avec la sonde oligonucléotidique FGPS518 (tableau 2).

La région amplifiée est une partie de la cassette construite spécifiquement pour obtenir la souche OS48.3 (CLERC-BARDIN et al., non publié).

Les comptes de calibration ont été dans tous les cas obtenus en utilisant l'ADN purifié de l'organisme cible.

2. RESULTATS

2.1 CHOIX DU TAMPON D'EXTRACTION:

20 sols différents ont été utilisés afin de déterminer le pH optimal du tampon d'extraction d'ADN. Pour tous les sols, le rendement en ADN augmente avec les pH croissants du tampon. Le rendement pour chaque pH (+/- sd), calculé comme le pourcentage de la valeur la plus haute pour chacun des sols, est le suivant: pH 6,0 : 31 +/- 13; pH 7,0: 43 +/- 16; pH 8,0: 60 +/- 14; pH 9,0: 82 +/- 12; pH 10,0: 98 +/- 3.

Pour 16 des 20 sols, le rendement le plus élevé a été obtenu à pH 10,0, alors que pour les quatre autres sols le plus haut rendement a

été obtenu à pH 9,0. Cependant, à pH 10,0, des quantités plus grandes de matériel humique ont été libérées, comparées à pH 9,0 (résultats non présentés). En conséquence, le pH 9,0 a été choisi pour toutes les expériences présentées ci-dessous.

5

2.2 EFFICACITE DES PROTOCOLES D'EXTRACTION D'ADN:

L'ADN total des organismes indigènes du sol a été extrait et quantifié de manière à évaluer l'efficacité de nombreux protocoles de lyse cellulaire *in situ*. Des échantillons des sols 1-6 (tableau 1) ont été traités selon les protocoles n°1 à 5 décrits dans la section Matériel et Méthodes (figure 1).

Après l'extraction d'ADN, les suspensions de sols ont été précipitées avec de l'isopropanol, et des fractions aliquotes des culots remis en suspension ont été analysées par électrophorèse sur gel, dans une première étape, afin d'estimer la qualité et la quantité de l'ADN libéré.

Cependant, la couleur de l'extrait d'ADN devenait de plus en plus sombre au fur et à mesure du nombre croissant d'étapes de lyse, du fait de la co-extraction de composés, tels que les acides humiques, avec l'ADN.

Certains de ces extraits bruts de couleur sombre ne migrent pas de la manière attendue dans les gels d'agarose.

En conséquence, les solutions d'ADN brut ont été purifiées (protocole B) avant quantification. Les électrophorèses sur gel des solutions purifiées obtenues après les différents traitements de lyse sont exemplifiées sur le sol n°3 (figure 2).

Une comparaison visuelle au rayonnement ultra-violet des intensités de l'ADN coloré a permis une estimation semi-quantitative de l'efficacité des traitements. De plus, la présence de profils de migration de tailles multiples de fragments (bandes discrètes) d'ADN et la disparition des fragments longs indique qu'une dégradation de l'ADN a eu lieu.

Aucun ADN n'a pu être extrait du sol argileux n°5.

Une quantification plus précise de l'ADN de tous les sols, extrait selon les protocoles n°1 à 5, a été réalisée par hybridation sur tache (Dot Blot) sans étape d'amplification PCR préalable et en utilisant une sonde oligonucléotidique complémentaire d'une séquence hautement conservée de la région d'ADNr 16S (sonde FGPS 431, tableau 2).

L'ADN a été détecté dans les extraits de tous les sols après chacune des différentes étapes de lyse, à l'exception du sol argileux n°5.

Les résultats concordent avec les estimations réalisées après gel d'électrophorèse.

- 10 Afin de comparer avec une méthode indépendante pour la quantification, l'ADN extrait selon le protocole n°2 (tous les sols sauf le sol n°5) a été également quantifié en utilisant une méthode colorimétrique de détection de l'ADN (RICHARD, 1974).

- 15 On a trouvé une bonne corrélation ($r = 0,88$) entre l'ADN quantifié en utilisant cette technique colorimétrique et les résultats obtenus par hybridation de type Dot Blot/radio-imagerie, confirmant l'hypothèse selon laquelle le nombre de copies moyen des bactéries du sol (mm) est de 7.

- 20 L'hybridation (Dot Blot) a montré que les quantités d'ADN extracellulaires, comme déterminé par extraction sans traitement de lyse (protocole n°1), allait de 4 µg/g pour le sol acide (n°6) à 36 µg/g pour le sol n°3 (tableau 3).

- 25 Le broyage du sol (protocole n°2) a augmenté les quantités d'ADN extrait à partir de tous les sols (p.ex. 26 µg/g de sol) pour le sol n°6 et 59 µg/g de sol (pour le sol n°3) (tableau 3; figure 2).

Pour les deux traitements de broyage (voir la section Matériel et Méthodes) la migration discrète d'ADN a été détectée sur les gels d'agarose, indiquant que les molécules d'ADN ont été partiellement dégradées (figure 2).

- 30 La taille des fragments d'ADN est comprise entre 20 et 0,2 kb. L'intensité de bande des fragments les plus petits est très faible, indiquant que la majeure partie des fragments ont une taille bien supérieure à 1 kb.

- 35 Le protocole n°3 comprend une étape d'homogénéisation dans un dispositif mixeur de type Ultraturax après l'addition du tampon

d'extraction aux échantillons de sol. Cette étape conduit à une augmentation des quantités d'ADN extrait, comme déterminé par hybridation sur tache (Dot Blot) pour deux des sols (le terreau sablonneux n°3 et le sol acide n°6), alors que les deux sols riches en matière organique (sols n°1 et n°2) ont conduit à l'obtention de quantités plus faibles d'ADN.

Les protocoles n°4a et n°4b ont permis d'évaluer l'influence de deux types de sonication sur les rendements en ADN à partir de sols préalablement broyés et homogénéisés .

La sonication n'a pas eu d'effet positif sur le rendement en ADN, comparé au protocole n°3, excepté pour le sol n°6. Toutefois, l'efficacité de lyse des deux types de sonicateur différent. Pour les sols n°2, 3 et 4, les quantités d'ADN extraites les plus grandes ont été obtenues en utilisant la micropointe de titane (tableau 3; figure 2), alors que pour les sols n°1 et n°6, le rendement en ADN était supérieur en utilisant le dispositif Cup Horn.

Des résultats contradictoires ont été également obtenus lorsque l'on a ajouté une étape de lyse enzymatique/chimique (protocoles n°5a et 5b) après l'étape de sonication: dans certains cas, les quantités d'ADN extraites ont été plus grandes que celles récupérées selon les protocoles n°4a et 4b, alors que dans d'autres cas les rendements étaient moindres (tableau 3).

2.3 COMPTAGE DIRECT DES MICRO-ORGANISMES:

Des comptes au microscope du nombre total de cellules bactériennes après coloration à l'acridine orange ont été réalisés pour tous les sols, avant et après broyage.

Avant broyage, le nombre de bactéries par gramme de poids sec du sol allait de $1,4 \times 10^9$ (+/- 0,4) dans le sol tropical n°5 à 10×10^9 (+/- 0,7) dans le sol provenant de la Côte Saint-André (sol n°3) (tableau 1).

Après broyage, les nombres de cellules ont été respectivement de 45, 74, 75, 54, 34 et 75% des valeurs initiales pour les sols n°1 à 6.

2.4 NUMERATION DES ACTINOMYCETES CULTIVABLES APPARTENANT A DIFFERENTS GENRES:

Une modification dans les populations d'actinomycètes dans le sol n°3 a été remarquée après les différents traitements de lyse (figure 3).

Par exemple, les colonies de *Streptomyces sp.* dominaient la flore viable d'actinomycètes lorsqu'aucun traitement de lyse n'est appliqué (protocole n°1), et représentaient 65% du nombre total de colonies identifiées. Après broyage, le pourcentage de colonies de *Streptomyces* a diminué pour atteindre 51%, alors que la proportion de colonies appartenant au genre *Micromonospora* a augmenté de 14% à 41%.

La lyse chimique/enzymatique (protocoles 5a et 5b) est apparue comme particulièrement efficace pour la lyse des streptomycètes. Lorsque tous les traitements de lyse ont été appliqués, y compris une lyse chimique/enzymatique (protocoles 5a et 5b), la microflore d'actinomycètes, qui comprenait encore plus de 10^8 CFU/g de sol, était dominée par les espèces appartenant au genre *Micromonospora*, alors qu'aucune ou très peu de colonies de *Streptomyces* ont été récupérées.

Les organismes appartenant aux genres tels que *Streptosporangium*, *Actinomadura*, *Microbispora*, *Dactylosporangium* et *Actinoplanes* sont apparus sur les plaques en faible nombre (2-8% du nombre total de colonies identifiées) après broyage, homogénéisation avec le dispositif Ultraturrax, et sonication, mais étaient généralement absents lorsque ces traitements étaient combinés avec une lyse chimique/enzymatique.

Le nombre total de bactéries cultivables restant après chaque traitement de lyse (protocoles 2 à 5) a été aussi recherché pour le sol n°4. Les résultats indiquent que le nombre de bactéries cultivables ne décroît pas avec l'intensité des traitements de lyse (environ 2×10^8 CFU/g de sol dans tous les cas, et également lorsqu'un traitement n'est appliqué, tel que selon le protocole n°1).

L'obtention de ces faibles valeurs de CFU est probablement due au fait que du sol sec a été utilisé et que seules les bactéries les

plus résistantes se sont multipliées sur les plaques. Le nombre d'actinomycètes formant colonies était généralement plus grand que celui des CFU total (toutes les bactéries) du fait qu'une étape de germination de spores, comprise dans le protocole de détection des actinomycètes, manquait lors du contrôle des bactéries totales.

2.5 RECUPERATION DE L'ADN DU PHAGE LAMBDA AJOUTE:

Le but de ces expériences était d'estimer de quelle manière des traitements de lyse successifs pouvaient affecter la récupération d'ADN nu, et si ces traitements successifs de lyse contribuaient à sa dégradation.

L'ADN pouvait être soit une fraction d'ADN extracellulaire libérée à partir d'organismes déjà morts, qui peuvent persister dans le sol pendant des mois (WARD et al., 1990), soit de l'ADN libéré à partir d'organismes lysés facilement pendant les premières étapes du traitement. Afin de simuler cette situation, de l'ADN de phage lambda digéré par HindIII a été ajouté, à diverses concentrations, aux sols avant et après broyage. En plus du broyage, une combinaison des autres traitements de lyse a été testée, y compris la sonication (dispositif Cup Horn, voir protocole n°4b) et des chocs thermiques (voir la section Matériel et Méthodes).

Après extraction, des fractions aliquotes qui devraient théoriquement contenir de 25 à 150 ng d'ADN de phage lambda ont été analysées par électrophorèse sur gel. Aucun fragment d'ADN spécifique du phage lambda n'a pu être observé lorsque l'ADN a été inoculé dans les échantillons de sol préalablement au broyage, indépendamment de la dose ou du type de sol.

Lorsque l'ADN a été ajouté après broyage, et extrait sans étape de traitement de lyse supplémentaire, les profils spécifiques d'ADN de phage lambda ont été détectés dans les extraits de quatre des cinq sols testés.

Dans tous ces cas, une relation directe de cause à effet a été obtenue entre la quantité d'ADN ajoutée et l'intensité des signaux sur les gels d'agarose. Les intensités des signaux étaient, cependant,

inférieures aux intensités de signaux attendues si on les compare à celles des standards moléculaires.

De plus, la bande à 23 kb était absente dans plusieurs cas, indiquant que les longs fragments étaient préférentiellement adsorbés
5 aux particules du sol, ou étaient plus sensibles à la dégradation, comparés aux fragments courts.

Aucune bande n'a été détectée dans les échantillons de sol tropical n°5 qui est caractérisé par une très haute teneur en argile (tableau 1).

10 Pour une quantification plus précise, la récupération d'ADN a été déterminée sur un dispositif d'imagerie par phosphorescence (phospho-imager) après hybridation en tache (Dot Blot). Selon cette technique, l'ADN a été détecté dans tous les échantillons, y compris ceux qui avaient été inoculés avant broyage, à l'exception du sol n°5
15 dans lequel aucun ADN n'a pu être détecté.

Dans tous les autres sols, la quantité d'ADN extrait augmente avec l'augmentation de taille de l'inoculum (figures 4a-d).

Cependant, les récupérations d'ADN de phage lambda étaient faibles. Lorsque le broyage était le seul traitement de lyse appliqué, les
20 récupérations étaient comprises entre 0,6 et 5,9% de l'ADN ajouté lorsque celui-ci était ajouté avant broyage, et de 3,6 à 24% de l'ADN ajouté lorsque ce dernier était ajouté après broyage. Les plus hauts niveaux de récupération ont été obtenus à partir du sol n°2.

L'électrophorèse sur gel de fractions aliquotes d'échantillons
25 traités par choc thermique et sonication n'a permis d'observer des bandes d'ADN dans aucun des échantillons, y compris l'essai dans lequel l'ADN avait été ajouté après broyage. Les expériences d'hybridation en tache (Dot Blot) ont confirmé ces résultats.

Les signaux d'hybridation obtenus à partir de suspensions de
30 sol qui ont été traitées par chocs thermiques et sonication ont été, tout au plus, faibles.

L'échantillon présentant la plus forte quantité d'ADN (15 µg d'ADN/g de poids sec du sol) était le seul pour lequel le signal obtenu était sensiblement différent du niveau du bruit de fond.

Aucune différence, (ou de faibles différences) n'a été observée entre les échantillons traités par choc thermique et ceux traités par chocs thermiques et sonication, indiquant que les chocs thermiques ont un effet préjudiciable sur l'ADN. Les récupérations les meilleures ont été observées pour le sol n°2, qui a la plus forte teneur en matière organique (tableau 1), alors qu'aucun ADN n'a été récupéré à partir du sol argileux n°5.

Des expériences additionnelles ont été réalisées avec des échantillons non broyés de sols n°4 et n°5, qui ont été ensemencés avec 20 et 50 µg d'ADN de phage lambda par gramme de sol.

Les échantillons ont été extraits immédiatement ou après une période d'incubation d'une heure à 28°C, puis les extraits d'ADN ont été purifiés et analysés par électrophorèse sur gel.

L'incubation du sol n°4 pendant une heure après l'inoculation n'a pas conduit à des profils qualitativement ou quantitativement différents de ceux obtenus sans incubation ou de ceux observés antérieurement lorsque l'ADN avait ajouté après broyage.

Ces résultats indiquent que la dégradation enzymatique par les nucléases du sol ne seraient pas impliquée dans le faible taux de récupération d'ADN. De plus, l'absence d'étape de broyage ne permet pas une augmentation de la récupération de l'ADN à partir du sol n°5, indiquant que les modifications de structure du sol dues au broyage n'augmentent pas significativement l'adsorption des acides nucléiques sur les colloïdes.

2.6 SATURATION DES SITES D'ADSORPTION AVEC L'ARN:

La plupart des profils obtenus sur les gels d'agarose ne diffèrent pas significativement des profils précédents dans lesquels le traitement d'ARN n'a pas été effectué.

Par exemple, aucune bande n'a été détectée à partir du sol riche en argile n°5, indépendamment des concentrations d'ARN et des concentrations d'ADN de phage lambda utilisées.

De plus, les bandes spécifiques d'ADN de phage lambda digérées par HindIII restaient indétectables dans le terreau sablonneux traité par l'ARN (sol n°4) lorsque l'ARN est ajouté avant le broyage.

- 5 L'intensité des bandes obtenues à partir d'échantillons ensemencés avec l'ADN après broyage augmente avec la concentration d'ARN, indiquant que le traitement pourrait avoir un effet positif.

Cependant, les résultats après hybridation et analyse par imagerie à phosphorescence n'ont pas confirmé les résultats de l'électrophorèse. Par exemple, l'effet positif du traitement d'ARN sur la
10 récupération d'ADN à partir du terreau argileux, lorsque l'ADN a été ajouté après broyage, n'apparaît pas clairement.

D'un autre côté, un effet positif de l'ARN a été trouvé pour le sol riche en argile (n°5) lorsque l'ADN a été ajouté après broyage.

- 15 Bien que les signaux d'hybridation pour les échantillons contrôle ne diffèrent pas des niveaux de bruit de fond, des quantités significatives d'ADN ont été libérées à partir des échantillons traités par l'ARN, et les signaux ont augmenté avec la quantité d'ADN ajoutée ainsi qu'avec la concentration d'ARN.

Cependant, même pour la plus forte concentration d'ARN (100
20 mg/g de poids de sol sec) le taux de récupération n'a jamais dépassé 3%.

2.7 PURIFICATION DES EXTRAITS BRUTS D'ADN:

- 25 Des quatre protocoles testés, la meilleure amplification des extraits d'ADN non dilués (1 µl d'extrait dans 50 µl de mélange PCR) a été observée après l'élution à travers des colonnes de type Microspin S400 suivie d'une élution à travers une colonne de type Elutip d, comme le montre l'électrophorèse sur gel des produits PCR.

30 L'ADN purifié par le système aqueux double phase (protocole C) a donné des quantités plus faibles de produits PCR après amplification à partir d'extrait d'ADN non dilué.

Aucun produit d'amplification n'a pu être obtenu à partir des extraits non dilués après amplification à la suite de la mise en oeuvre
35 des protocoles A ou B. En conséquence, le protocole B (voir section

Matériels et Méthodes) a été utilisé pour toutes les expériences dans lesquelles les amplifications PCR et/ou les hybridations sur tache (Dot Blot) ont été réalisées.

5 2.8 QUANTIFICATION PAR PCR ET HYBRIDATION:

La première étape était de déterminer si les quantités de produit PCR étaient proportionnelles au nombre de molécules d'ADN cibles initialement présentes dans le tube réactionnel. De l'ADN de
10 *Streptosporangium fragile* a été utilisé comme cible (voir section Matériels et Méthodes).

Les amorces utilisées ont été les amorces FGPS122 et FGPS350 (tableau 2). L'électrophorèse sur gel des produits PCR a montré que l'intensité de bande augmente avec l'accroissement de la
15 concentration des cibles. Les produits PCR ont été hybridés avec la sonde oligonucléotidique FGPS643 (tableau 2), et les signaux ont été quantifiés par imagerie par phosphorescence (phospho-imaging).

On a trouvé une bonne corrélation ($r^2 = 0,98$) entre le \log [nombre de cibles] et le \log [intensité du signal d'hybridation].

On a ensuite recherché si l'efficacité de l'amplification PCR était affectée par les acides humiques et l'ADN non cible. Lorsqu'on l'analyse
20 par électrophorèse sur gel, l'intensité accrue des bandes des produits PCR, correspondant aux différentes quantités d'ADN cible, était conservée lorsque l'amplification était réalisée avec des solutions d'ADN
25 auxquelles on avait ajouté des extraits de sol traités à la DNase, contenant des acides humiques à des concentrations allant jusqu'à 8ng dans le mélange PCR d'un volume de 50 μ l.

Avec 20 ng d'acide humique dans le mélange PCR, les bandes correspondant aux faibles niveaux d'ADN cible ont disparu, et à des
30 concentrations d'acide humique de 80 ng et à des concentrations supérieures, aucune bande n'était visible.

Les quantités variées d'ADN cible de *S.fragile* ont permis de fournir les quantités attendues de produit PCR lorsque, avant amplification, l'ADN de *S. fragile* a été mélangé avec de l'ADN de
35 *Streptomyces hygroscopicus* et ajouté au mélange PCR de 50 μ l dans

une gamme de 100 pg à 1µg afin de simuler l'ADN non-cible libéré à partir de la microflore du sol.

2.9 QUANTIFICATION DES ACTINOMYCETES INDIGENE DU SOL APRES DIFFERENTS TRAITEMENTS DE LYSE:

On a appliqué le protocole de purification D suivi d'une amplification par PCR comme décrit ci-dessus afin de quantifier les actinomycètes appartenant au genre *Streptosporangium* dans le sol n°3 après extraction conformément aux protocoles n°1, 2, 3, 5a et 5b (figure 5).

Après broyage, (protocole n°2) la quantité d'ADN cible provenant de cet actinomycète a été estimée par hybridation (Dot Blot) et radio-imagerie comme étant de 2,5 +/- 1,3 ng /g de poids de sol sec.

Si l'on postule que le contenu en ADN est de 10 fg par cellule, comme pour *Streptomyces* (Gladek et al. 1984), cette valeur correspond à approximativement $2,5 \times 10^5$ génomes. Des valeurs similaires ont été obtenues après les autres traitements de lyse (respectivement 2,6 +/-1,1 et 1,8 +/- 1,3 ng d'ADN/g de sol sec en utilisant respectivement les protocoles 3 et 4b).

2.10 EFFICACITE DE LA RECUPERATION D'ADN A PARTIR DE SOLS PREALABLEMENT INOCULES AVEC DES BACTERIES:

Trois sols (n°2, 3 et 5) ont été inoculés avec des spores ou des hyphae de *Streptomyces lividans* à différentes concentrations (voir section Matériel et Méthodes). Les quantités de mycélium ajoutées au sol (figure 6b) correspondent au nombre de spores inoculées dans le milieu de germination. Approximativement 50% de ces spores ont germé. Le nombre exact de cellules dans les hyphae des spores germées n'a pas été déterminé. En conséquence, les quantités de spores et de mycélium ensemencées dans les sols ne sont pas directement comparables.

Pour chaque échantillon de sol, le protocole d'extraction n°6, la méthode de purification D, et l'amplification PCR combinée avec l'hybridation sur tache (Dot Blot) et l'imagerie par phosphorescence (phospho-imaging) ont été utilisés pour dénombrer les ADNs cibles

spécifiques qui avaient été libérés. L'ADN extrait peut être clairement distingué du bruit de fond seulement lorsque le nombre de spores ajoutées dépasse 10^5 pour les sols n°3 et n°5 et 10^7 pour le sol n°2 (figure 6a).

5 Lorsque le mycélium est ajouté, l'ADN extrait peut être détecté au-delà d'une quantité correspondant à 10^3 spores/g de sol pour les sols n°2 et n°3, et au-delà de 10^7 spores/g pour le sol n°5 (figure b).

Au-dessus du niveau de détection, le signal d'hybridation augmente avec des quantités croissantes des cellules inoculées.

10 Pour l'inoculum de spores, une augmentation de 100 fois dans le nombre de cellulesensemencées conduit à une augmentation de presque 100 du rendement d'ADN. Cette augmentation est clairement inférieure lorsque les hyphae sont inoculées, particulièrement dans les sols n°2 et n°3 (figure 6).

15 Au contraire, les résultats obtenus lorsque l'ADN de phage lambda a été utilisé comme inoculum, l'ADN a également été récupéré à partir du sol riche en argile (n°5) lorsque les cellules bactériennes ont été utilisées comme inoculum. Cependant, pour ce dernier aussi, le traitement par l'ARN a augmenté la récupération d'ADN de
20 *Streptomyces* à partir de ce sol à la fois pour les spores et le mycélium (figure 6).

Le fait d'ensemencer des sols avec des cellules végétatives de *Bacillus anthracis* a fourni des taux de récupération similaires à ceux obtenus pour *Streptomyces*.

25 De plus, les taux de récupération d'ADN à partir du sol n°5 ont augmenté après traitement par l'ARN également pour cet inoculum.

**Exemple 2 : Construction d'une banque d'ADN de faible poids moléculaire (<10 kb) à partir d'un sol contaminé par du lindane :
30 clonage et expression du gène *linA***

Cet exemple décrit la construction d'une librairie d'ADN du sol dans *E. coli*. Il permet de démontrer le clonage et l'expression de gènes de petite taille issus d'une microflore non cultivable .

Le lindane est un pesticide organochloré, récalcitrant à la dégradation et persistant dans l'environnement. En aérobie, sa biodégradation est catalysée par une déhydrochlorinase, codée par le gène *linA*, permettant de transformer le lindane en 1,2,4-trichlorobenzène. Le gène *linA* n'a été identifié que parmi deux souches isolées du sol : *Sphingomonas paucimobilis*, isolé au Japon (Seeno et Wada 1989, Imai *et al* 1991, Nagata *et al* 1993) et *Rhodanobacter lindaniclasticus* isolé en France (Thomas *et al* 1996, Nalin *et al* 1999).

Pourtant le potentiel de dégradation du lindane, mis en évidence par dosage des ions chlorures libérés et amplification par PCR du gène *linA* à partir de sols ayant été en contact ou non avec du lindane, semble être répandu plus largement dans l'environnement (Biesiekierska-Galguen, 1997).

15 **1. Extraction directe d'ADN de sol**

Les sols secs sont broyé pendant 10 minutes dans un broyeur à force centrifuge Restch équipé 6 billes de tungstène. 10 grammes de sol broyé sont mis en suspension dans 50 ml de tampon TENP pH 9 (Tris 20 50 mM, EDTA 20 mM, NaCl 100 mM, polyvinylpolypyrrolidone 1% w/v), et homogénéisés au vortex pendant 10 min.

Après centrifugation de 5 minutes, 4000 g à 4°C, le surnageant est précipité à l'acétate de sodium (3M, pH 5.2) et à l'isopropanol, pour être repris dans du tampon TE stérile (Tris 10 mM, EDTA 1 mM, pH 8.0). L'ADN extrait est ensuite purifié sur colonne de tamisage moléculaire S400 (Pharmacia) et sur colonne échangeuse d'ions Elutip d (Schleicher et Schuell), selon les instructions des fabricants, puis conservé dans du TE.

30 **2. Construction de la banque d'ADN extrait du sol dans le vecteur pBluescript SK-**

Le vecteur pBluescript SK- et l'ADN extrait du sol sont chacun digérés par les enzymes *HindIII* et *BamHI* (Roche), à raison de 10 unités d'enzymes pour 1 µg d'ADN (incubation 2 heures à 37°C). Les ADN sont

ensuite ligués par action de la T4 DNA ligase (Roche), une nuit à 15°C, à raison d'une unité d'enzyme pour 300 ng d'ADN (environ 200 ng d'ADN insert et 100 ng de vecteur digéré). Les cellules d'*Escherichia coli* électrocompétentes, ElectroMAX DH10B™ (Gibco BRL) sont transformées par le mélange de ligation (2 µl) par électroporation (25 µF, 200 et 500 Ω, 2,5 kV) (Biorad Gene Pulser).

Après une heure d'incubation dans le milieu LB, les cellules transformées sont diluées de façon à obtenir environ 100 colonies par boîte puis sont étalées sur milieu LB (10 g/l Tryptone, 5 g/l levure, 5 g/ NaCl) additionné d'Ampiciline (100 mg/l), de γ-HCH (500 mg/l), de X-gal 60 mg/l (5-bromo-4-chloro-3-indolyl-α-D-galactoside), et d'IPTG 40 mg/l (isopropylthio-β-D-galactoside), et incubées une nuit à 37°C. Le γ-hexachlorocyclohexane (Merck-Schuchardt) étant insoluble dans l'eau, une solution à 50 g/l est préparée dans du DMSO (dimethyl sulfoxyde) (Sigma).

Une banque de 10 000 clones a ainsi été obtenue.

3.Clonage et expression du gène *linA*

Le criblage de la banque s'effectue par visualisation d'un halo de dégradation du lindane autour de la colonie (le lindane précipitant dans les milieux de culture). Sur 10 000 clones criblés, 35 présentaient ainsi une activité de dégradation du lindane. La présence du gène *linA* chez ces clones a pu être confirmée par PCR grâce à des amorces spécifiques, décrites par Thomas et al (1996). Des digestions réalisées sur les inserts ainsi que sur les produits d'amplification ont montré des profils identiques entre tous les clones criblés et le témoin de référence, *R. lindaniclasticus*. Les clones portant le gène *linA* présentaient également un insert de même taille (environ 4 kb).

Il a ainsi pu être démontré que l'ADN du sol pouvait être cloné et exprimé chez un hôte hétérologue : *E. coli*, et que des gènes issus d'une microflore difficilement cultivable pouvaient être exprimés. Des banques réalisées à partir de digestion partielle d'ADN extrait du sol par des enzymes de restriction telles que *Sau3AI* sont donc aussi envisageables.

EXEMPLE 3:

Procédé de préparation d'une collection d'acides nucléiques à partir d'un échantillon de sol, comprenant une étape d'extraction indirecte de l'ADN.

1. MATERIEL ET METHODES.**1.1 Extraction de la fraction bactérienne du sol.**

5g de sol sont dispersés dans 50 ml de NaCl 0.8% stérile, par broyage au Waring Blender pendant 3 x 1 minute, avec refroidissement dans la glace entre chaque broyage. les cellules bactériennes sont alors séparées des particules du sol par centrifugation sur un coussin de densité de Nycodenz (Nycomed Pharma AS, Oslo, Norvège). Dans un tube à centrifugation, 11,6 ml d'une solution de Nycodenz de densité de 1.3 g.ml⁻¹ (8g de Nycodenz suspendu dans 10 ml d'eau stérile) sont placés en dessous de 25 ml de la suspension de sol précédemment obtenue. Après centrifugation à 10.000 g dans un rotor à godets mobiles (rotor TST 28.38, Kontron) pendant 40 minutes à 4°C, l'anneau cellulaire, se situant à l'interphase de la phase aqueuse et de la phase Nycodenz, est prélevé, lavé dans 25 ml d'eau stérile et centrifugé à 10.000 g pendant 20 minutes. Le culot cellulaire est ensuite repris dans une solution Tris 10 mM; EDTA 100 mMn pH 8.0.

Préalablement à la dispersion du sol au Waring Blender, une étape d'enrichissement du sol dans une solution d'extrait de levure peut être incluse afin de permettre notamment la germination des spores bactériennes du sol. 5 g de sol sont alors incubés dans 50 ml d'une solution stérile de NaCl 0.8% - extrait de levure 6%, pendant 30 minutes à 40°C. L'extrait de levure est éliminé par centrifugation à 5000 rpm pendant 10 minutes afin d'éviter la formation de mousse durant le broyage.

1.2 Lyse des cellules bactériennes du sol.

- Lyse des cellules en milieu liquide et purification sur gradient de chlorure de césium.

Les cellules sont lysées dans une solution Tris 10 mM, EDTA 100 mM, pH 8.0 contenant 5 mg.ml⁻¹ de lysozyme et 0.5 mg.ml⁻¹ d'achromopeptidase pendant 1 heure à 37°C . Une solution de lauryl sarcosyl (1% final) et de protéinase K (2 mg.ml⁻¹) est ensuite ajoutée et incubée à 37°C pendant 30 minutes. La solution d'ADN est alors purifiée sur un gradient de densité de chlorure de césium par centrifugation à 35 000 rpm pendant 36 heures sur un rotor Kontron 65.13. Le gradient de chlorure de césium employé est un gradient à 1g/ml de CsCl, possédant un indice de réfraction de 1,3860 (Sambrook et al., 1989).

- Lyse des cellules après inclusion dans un bloc d'agarose.

Les cellules sont mélangées à un volume égal d'agarose à 1.5% (poids/volume) Seaplaque (Agarose Seaplaque FMC Products. TEBU, Le Perrat en Yvelines, France). à bas point de fusion et coulées dans un bloc de 100 µl. Les blocs sont ensuite incubés dans une solution de lyse : EDTA 250 mM, saccharose 10.3%, lysozyme 5 mg.ml⁻¹ et achromopeptidase 0.5 mg.ml⁻¹ à 37°C pendant 3 heures. Les blocs sont alors lavés dans une solution de Tris 10 mM - EDTA 500 mM et incubés une nuit à 37°C dans de l'EDTA 500 mM contenant 1 mg.ml⁻¹ de protéinase K et du lauryl sarcosyl 1%. Après plusieurs lavages dans du Tris-EDTA, les blocs sont conservés dans de l'EDTA 500 mM.

La qualité des ADN ainsi extraits est contrôlée par électrophorèse en champs pulsés.

La quantité d'ADN extrait a été évaluée sur gel d'électrophorèse par rapport à une gamme étalon d'ADN de thymus de veau.

1.3 Caractérisation moléculaire de l'ADN extrait du sol.

Les ADN extraits du sol sont caractérisés par hybridation PCR, méthode qui consiste à amplifier dans un premier temps les ADNs à l'aide d'amorces situées sur des régions universellement conservées du gène de l'ARNr16S, puis à hybrider les ADNs amplifiés avec différentes sondes oligonucléotidiques de spécificité connue (tableau 4), dans le but

de quantifier l'intensité du signal d'hybridation par rapport à une gamme étalon externe d'ADN génomique.

Les ADN extraits du sol ainsi que les ADN génomiques extraits de cultures pures sont amplifiés avec les amorces FGPS 612-669 (tableau 1) dans les conditions standard d'amplification par PCR. Les produits d'amplification sont ensuite dénaturés par un volume égal de NaOH 1N, déposés sur une membrane de Nylon (GeneScreen Plus, Life Science Products) et hybridés avec une sonde oligonucléotidique marquée à son extrémité par du $g^{32}P$ ATP par action de la T4 polynucléotide kinase. Après préhybridation de la membrane dans une solution de 20 ml contenant 6 ml de SSC 20X, 1 ml de solution de Denhardt, 1 ml de SDS 10% et 5 mg d'ADN hétérologue de sperme de saumon, les hybridations sont conduites durant une nuit à la température définie par la sonde. Les membranes sont lavées deux fois dans du SSC 2X pendant 5 minutes à température ambiante, puis une fois dans du SSC 2X SDS 0,1% et une seconde fois dans du SSC 1X, SDS 0,1% pendant 30 minutes à la température d'hybridation. Les signaux d'hybridation sont quantifiés à l'aide du logiciel Molecular Analyst (Biorad, Ivry sur Seine, France) et les quantités d'ADN sont estimées par interpolation des courbes étalons obtenues à partir des ADN génomiques.

2. RESULTATS ET DISCUSSION

2.1 Extraction et lyse de la fraction bactérienne du sol.

La séparation des cellules microbiennes des particules du sol, préalablement à l'extraction de l'ADN, est une alternative présentant de nombreux avantages par rapport aux méthodes d'extraction directe de l'ADN dans le sol. En effet, l'extraction de la fraction microbienne limite la contamination de l'extrait d'ADN par de l'ADN extracellulaire présent librement dans le sol ou par de l'ADN d'origine eucaryote. Mais surtout, l'ADN extrait de la fraction microbienne du sol présente des fragments de plus longue taille et une meilleure intégrité que l'ADN extrait par lyse directe JACOBSON et RASMUSSEN (1992). De plus, la séparation des

particules de sol permet d'éviter une contamination de l'extrait d'ADN par des composés humiques et phénoliques, composés pouvant, par la suite, nuire gravement aux efficacités de clonage.

Une des étapes déterminantes pour l'extraction des cellules du sol est la dispersion de l'échantillon de sol afin de dissocier les cellules adhérent à la surface ou à l'intérieur des agrégats de particules de sol. Trois cycles de broyage successifs d'une minute chacun permettent d'obtenir une meilleure efficacité d'extraction des cellules ainsi qu'une plus grande quantité d'ADN récupéré, par rapport à un unique cycle de broyage d'une minute 30.

Le tableau 5 rapporte les efficacités d'extraction obtenues après centrifugation sur gradient de Nycodenz, sur la microflore totale viable (dénombree par microscopie après coloration à l'acridine orange), sur la microflore totale cultivable (dénombree sur milieu solide Trypticase-Soja 10%), et sur la microflore d'actinomycètes cultivables sur milieu HV agar (après incubation à 40°C dans une solution d'extrait de levure 6% -SDS 0,05% afin de provoquer la germination des spores). D'autre part, l'ADN extrait a été quantifié soit après une lyse des cellules en milieu liquide (sans purification sur gradient de chlorure de césium) soit après une lyse des cellules incluses dans un bloc d'agarose (après digestion de l'agarose par une b-agarase).

Les résultats montrent que plus de 14% de la microflore tellurique totale est récupéré par cette méthode (soit 2×10^8 cellules par gramme de sol), et que la microflore totale cultivable ne représente qu'à peine 2% de la population microbienne totale.

D'autre part, la quantité d'ADN extrait des cellules est de 330 ng par gramme de sol sec. En estimant le contenu d'ADN par cellule microbienne du sol entre 1.6 et 2.4 fg, et compte tenu de la quantité de cellules extraites (2×10^8 cellules par gramme de sol), on peut estimer que la quasi-totalité des cellules ont été lysées et qu'ainsi la lyse n'apporte pas d'important biais à cette approche.

Les électrophorèses en champs pulsés ont montré que l'ADN du sol extrait après gradient de Nycodenz et de CsCl pouvait atteindre une taille de 150 kb et que la lyse en bloc d'agarose permettait d'extraire des fragments supérieurs à 600kb.

Ces résultats confirment l'intérêt de cette approche indépendante de la culture pour la construction de banques d'ADN de l'environnement, en se présentant comme une alternative aux méthodes directes d'extraction d'ADN.

5

2.2 Caractérisation moléculaire de l'ADN extrait du sol.

Le but de la caractérisation moléculaire de l'ADN extrait du sol est d'obtenir des profils représentant les proportions des différents taxons bactériens présents dans l'extrait d'ADN. Il s'agissait également de connaître les biais d'extraction induits par la séparation préalable de la réaction cellulaire du sol, en comparaison avec une méthode d'extraction directe faite de visualisation directe de la diversité microbienne présente dans les sols. En effet, peu d'informations ont été rassemblées sur l'extraction des cellules sur gradient de Nycodenz en fonction de leur structure morphologique (diamètre des cellules, formes filamenteuses ou sporulées).

Les méthodes jusqu'ici en place étaient basées sur des:

- hybridations quantitatives utilisant des sondes oligonucléotidiques spécifiques à différents groupes bactériens, appliqués directement d'ADN extrait de l'environnement. Malheureusement, cette approche n'est pas très sensible et ne permet pas de détecter des genres ou des groupes taxonomiques présents en faible abondance AMANN (1995).
- PCR quantitatives telles que la MPN-PCR (Most Probable Number) SYKES et al. (1992) ou la PCR quantitative par compétition DIVIACCO et al. (1993). Les inconvénients respectifs de chacune de ces approches sont (i) la lourdeur d'utilisation du fait de la multiplication des dilutions et des répétitions qui rend la technique inappropriée pour un grand nombre d'échantillons ou de couples d'amorces, et (ii) la nécessité de construire un compétiteur spécifique à l'ADN cible et n'induisant pas de biais dans la compétition.

La méthode mise en place selon la présente invention consiste à amplifier universellement un fragment de 700 pb à l'intérieur de la séquence d'ADNr 16S, à hybrider cet amplifiat avec une sonde

oligonucléotidique de spécificité variable (au niveau du règne, de l'ordre, de la sous classe ou du genre) et à comparer l'intensité d'hybridation de l'échantillon par rapport à une gamme étalon externe. L'amplification préalable à l'hybridation permet de quantifier des genres ou des espèces de micro-organismes peu abondants. De plus, l'amplification par des amorces universelles permet, lors de l'hybridation, d'utiliser une large série de sondes oligonucléotidiques. Elle permet de comparer entre eux différents modes de lyse (extraction directe ou indirecte) sur des groupes taxonomiques bien définis.

10 Les résultats sont rassemblés dans le tableau 6.

Ils montrent des profils similaires entre les deux méthodes d'extraction (directe et indirecte). Ainsi, il apparaît que l'extraction préalable de la fraction microbienne tellurique n'introduit pas de réels biais parmi les taxons testés. La seule différence significative entre les deux approches d'extraction semblerait être la plus grande abondance de séquences d'ADNr appartenant aux γ protéobactéries dans l'extrait par la méthode d'extraction indirecte.

De plus, un effet significatif de l'incubation de l'échantillon de sol dans une solution d'extrait de levure est observé sur les populations sporulées du sol (Gram⁺, bas pourcentage de GC et Actinomycètes). Cette étape provoque la germination des spores, et permet d'une part certainement une meilleure récupération de ce type de cellules et d'autre part une plus grande efficacité de la lyse sur des cellules en germination.

Cette approche permet une analyse semi-quantitative, ciblée sur les principaux taxons définis à partir de micro-organismes cultivés et habituellement retrouvés dans les sols. Seuls des outils moléculaires permettent d'estimer l'importance des différents taxons, les méthodes de mise en culture étant trop restrictives et dépendantes de la spécificité du milieu utilisé.

30 Les résultats montrent qu'une grande part de la population microbienne n'est pas représentée dans les groupes phylogénétiques décrits, mettant ainsi en évidence l'existence de nouveaux groupes composés de micro-organismes non cultivés jusqu'à présent, ou non cultivables.

Ainsi, de nouvelles sondes peuvent être définies à partir de séquences déterminées à partir d'ADN extrait du sol (nouveaux phylums composés de micro-organismes non cultivés, LUDWIG et al. (1997) afin d'obtenir une image plus exacte de la composition de l'extrait d'ADN.

5

Exemple 4 : - CONSTRUCTION DU COSMIDE POS 7001

Caractéristiques de POS 7001:

- Réplicatif chez *E. coli*
10 Intégratif chez *Streptomyces*
Sélectionnable chez *E. coli* AmpR, HygroR et *Streptomyces* HygroR
- Les propriétés du cosmide permettent d'insérer de grands fragments d'ADN entre 30 et 40kb.
- 15 Il comprend
- 1 - Le promoteur inductible *tipA* de *Streptomyces lividans*
 - 2 - Le système d'intégration spécifique de l'élément pSAM2
 - 3 - Le gène de résistance à l'hygromycine
 - 4- le cosmide pWED1, dérivé de pWED15

20

1) - Le promoteur inductible du gène *tip A* de *S. lividans*

- Le gène *tipA* code une protéine de 19 KD dont la transcription est induite par l'antibiotique thiostrepton ou nosiheptide. Le promoteur de
- 25 *tipA* est bien régulé: induction en phase exponentielle et en phase stationnaire (200X) Murakami T, Holt TG, Thompson CJ. J. Bacteriol 1989 ;171 :1459-66

30

2) - Le gène de résistance à l'hygromycine

- Hygromycine: antibiotique produit par *S. hygroscopicus*
- Le gène de résistance code une phosphotransférase (*hph*)
- Le gène utilisé provient d'une cassette construite par Blondelet et al dans laquelle le gène *hyg* est sous contrôle de son propre promoteur

et du promoteur plac inductible par l'IPTG Blondelet-Rouault et al ;
Gene 1997 ;190 :315-7

3) - Le système d'intégration site-spécifique

L'élément pSAM2 s'intègre dans le chromosome par un mécanisme d'intégration site-spécifique. La recombinaison a lieu entre deux séquences identiques de 58 pb présentes sur le plasmide (*attP*) et sur le chromosome (*attB*).

Le gène *int*, situé à proximité du site *attP*, est impliqué dans l'intégration site-spécifique de pSAM2, et son produit présente des similitudes avec les intégrases des bactériophages tempérés d'entérobactéries. Il a été démontré qu'un fragment de pSAM2 ne contenant que le site d'attachement *attP* ainsi que le gène *int* était capable de s'intégrer de la même manière que l'élément entier. Voir brevet français n°88 06638 du 18/05/1988, ainsi que Raynal A et al. Mol Microbiol 1998 28 :333-42).

4) - Construction du cosmide pOS700I

Etape 1/ Le promoteur TipA a été isolé du plasmide pPM927 (Smokvina et al. Gene 1990; 94:53-9) sur un fragment HindIII-BamHI de 700 paires de bases et cloné dans le vecteur pUC18 (Yannish-Perron et al., 1985) digéré par HindIII/BamHI

Etape 2/ Ce fragment HindIII-BamHI a ultérieurement été transféré de pUC18 à pUC19 (Yannish-Perron et al., 1985).

Etape 3/ Un insert BamHI-BamHI de 1500 paires de bases portant le gène *int* et le site *attP* de pSAM2 a été isolé du plasmide pOSint1, représenté à la Figure 8, (Raynal A et al. Mol Microbiol 1998 28 :333-42) et cloné au site BamHI du vecteur précédent (pUC19/TipA), dans l'orientation permettant de mettre le gène *int* sous contrôle du promoteur TipA.

Etape 4/ Le site BamHI situé en 5' du gène int a été supprimé par digestion partielle BamHI puis traitement par l'enzyme Klenow. Un fragment HindIII-BamHI portant TipA-int-attP a ainsi été isolé de pUC19 et transféré dans pBR322 HindIII/BamHI.

5

Etape 5/ La cassette Hygromycine isolée de pH45 Ω hyg (Blondelet-Rouault et al., 1997) sur un fragment HindIII-HindIII a été clonée au site HindIII situé en amont du promoteur TipA.

10 **Etape 6/** Le site HindIII situé entre la cassette Ω Hyg et le promoteur TipA a été supprimé par traitement Klenow après digestion partielle HindIII.

15 **Etape 7/** Le plasmide obtenu à l'issue de l'étape précédente permet d'isoler un fragment unique HindIII-BamHI, portant tous les éléments Ω Hyg/TipA/int attP, qui a été cloné après traitement Klenow au site EcoRV du cosmide pWED1. Le cosmide pWED1, représenté à la Figure 9, dérive du cosmide pWE15, représenté à la Figure 10 (Wahl GM, et al. Proc Natl Acad Sci U S A 1987 84:2160-4) par délétion d'un fragment
20 HpaI-HpaI portant le gène Neomycine et l'origine SV40.

Une carte du vecteur pOS 700I est représentée à la Figure 11.

25 **Exemple 5 : Construction de plusieurs cosmides conjugatifs et intégratifs chez *Streptomyces*, les vecteur pOSV 303, pOSV306 et pOSV307**

5.1 Construction du vecteur pOSV303.

30 Etant donné que l'emballage sélectionne les clones ayant une taille supérieure à 30kb, seuls 10 à 15% des clones ne contiennent pas d'insert, il n'est donc pas vraiment nécessaire d'avoir un système de sélection des recombinants, ce qui permet de construire un vecteur plus petit.

35

Construction:**Etape 1 : le vecteur pOSV001**

- Clonage d'un fragment PstI-PstI de 800 paires de bases portant l'origine de transfert OriT du réplicon RK2 (Guiney et al., 1983), dans le plasmide pUC19 ouvert par PstI. Cette étape de clonage permet d'obtenir un vecteur transférable de *E. coli* à *Streptomyces* par conjugaison.

La carte du vecteur pOSV 001 est représentée à la Figure 17.

10 Etape 2 : le vecteur pOSV002

Insertion du marqueur Hygromycine (cassette Ω hyg), et sélectionnable chez *Streptomyces*, de sorte que le gène conférant la résistance à l'hygromycine soit transféré en dernier ce qui permet de s'assurer du transfert complet du BAC avec l'insert d'ADN du sol.

- 15 Clonage de la cassette Hygromycine isolée de pHP45 Ω hyg sur un fragment HindIII-HindIII portant le gène de résistance à l'Hygromycine.. Ce fragment est cloné au site PstI (position 201) du vecteur pOSV001. Ce site PstI a été choisi, compte tenu du sens du transfert, pour que le marqueur Hygro soit le dernier transféré lors de la conjugaison. Les extrémités PstI et HindIII sont rendues compatibles après traitement par le fragment Klenow de l'ADN polymérase permettant de générer des "bouts francs". L'orientation du fragment Ω hyg est déterminée en fin de construction.

La carte du vecteur pOSV002 est représentée à la Figure 18.

25

Etape 3 : le vecteur pOSV010

Le fragment XbaI-HindIII isolé du plasmide pOSV002 et contenant le marqueur de résistance à l'hygromycine et l'origine de transfert est cloné dans le plasmide pOSint1 digéré par XbaI et HindIII. L'orientation des sites est telle que le marqueur hygromycine sera toujours transféré en dernier.

30

Le plasmide pOSint1, représenté à la Figure 8, a été décrit dans l'article de Raynal et al.(Raynal A et al. Mol Microbiol 1998 28 :333-42).

- 35 Cette construction permet l'expression de l'intégrase chez *E. coli* et chez *Streptomyces*.

Etape 4 : insertion du site "cos"

Le principe est d'insérer un site "cos" dans le plasmide pOSV010 permettant l'empaquetage dans le plasmide pOSV010, représenté à la Figure 12.

L'obtention du fragment "cos" est représentée à la Figure 13.

Ce fragment est obtenu par PCR. A partir d'un fragment portant les extrémités cohésives (cos) de λ (bactériophage lambda ou cosmide pHc79), une amplification par PCR est réalisée à l'aide des oligonucléotides correspondant aux séquences -50/+130 par rapport au site cos. Ces oligonucléotides contiennent en outre les sites de clonage NsiI, compatible PstI, XhoI, compatible Sall, EcoRV, "bout franc".

L'addition des sites rares SmaI et PaeI permet d'isoier et/ou de cartographier l'insert cloné.

Le fragment PCR est borné par un site PstI à l'extrémité 5' et par un site HincII à l'extrémité 3', permettant le clonage dans le vecteur pOSV010 (Figure 12) préalablement digéré par les enzymes NsiI et EcoRV, provoquant la délétion du répresseur lacIq.

La carte du vecteur pOSV303 est représentée sur la Figure 14. Le vecteur pOSV303, contient des sites de clonage tels que le site NsiI, compatible PstI, le site XhoI, compatible Sall ou encore le site EcoRV pour l'obtention de "bouts francs".

5.2 Construction du vecteur pOSV306**Etape 1: Construction du vecteur pOSV308.**

Le vecteur pOSV308 a été construit selon le procédé illustré à la figure 27. Un fragment de 643 pb contenant la région cos a été amplifié à l'aide du couple d'amorces de séquences SEQ ID N°107 et SEQ ID N°108 à partir du vecteur cosmide pHc79 décrit par HOHM B and COLLINS (1980).

Ce fragment nucléotidique amplifié a été cloné directement dans le vecteur pGEMT-easy commercialisé par la Société PROMEGA, comme illustré à la figure 27 afin de produire le vecteur pOSV308.

5 **Etape 2: Construction du vecteur pOSV306.**

Le vecteur pOSV010 a été construit comme décrit à l'étape 3 de construction du vecteur pOSV303, comme décrit au paragraphe 5.1 du présent exemple.

- 10 Le vecteur pOSV10 a été digéré par les enzymes EcoRV et NsiI afin d'exciser un fragment de 7874 pb qui a été ultérieurement purifié, comme cela est illustré à la figure 28.

- Puis, le vecteur pOSV308 obtenu à l'étape 1) ci-dessus a été soumis à une digestion par les enzymes EcoRV et PstI afin d'exciser un
15 fragment de 617 pb, qui a été ultérieurement purifié.

 Puis, le fragment cos de 617 pb obtenu à partir du vecteur pOSV308 a été intégré par ligation dans le vecteur pOSV10, afin d'obtenir le vecteur pOSV306, comme cela est illustré à la figure 28.

20 **5.3 Construction du vecteur pOSV307.**

Le cosmide pOSV307 contient toujours le gène LacIq, afin d'améliorer la stabilité du cosmide dans *Streptomyces*, par exemple dans la souche S17-1 de *Streptomyces*.

- 25 Afin de construire le vecteur pOSV307, on a soumis le vecteur pOSV010 à une digestion par l'enzyme PvuII, pour obtenir un fragment de 8761 pb qui a été purifié, puis déphosphorylé.

- Ensuite, le vecteur pOSV308, tel qu'obtenu comme décrit à l'étape 1) du paragraphe 5.2 ci-dessus, a été digéré par l'enzyme EcoRI
30 afin d'obtenir un fragment de 663 pb, qui a été ensuite purifié et traité par l'enzyme de Klenow.

 Le fragment nucléotidique ainsi traité a été intégré dans le vecteur pOSV010 après ligation afin d'obtenir le vecteur pOSV307, comme illustré à la figure 29.

Exemple 6 : - Construction du cosmide réplcatif navette *E. coli*-*Streptomyces* pOS700R.

Les fragments du plasmide pEI16 (Voff et al., 1996) représenté
5 à la Figure 15 ont été isolés et traités par Klenow. Ces fragments
contiennent les séquences nécessaires à la replication et à la stabilité
provenant du plasmide SCP2.

Ces deux fragment sont insérés séparément dans le site
EcoRV du cosmide pWED1 conduisant à 2 clones différents.

10 La cassette Hygromycine isolée de PHP45Ωhyg sur un
fragment HindIII-HindIII a été clonée au site HindIII des cosmides
pWED1 contenant l'insert ScP2 sous forme de fragments PstI-EcoRI ou
XbaI. Elle confère une résistance à l'Hygromycine sélectionnable à la
fois chez *E. coli* et chez *Streptomyces*.

15 Transformation de *S. lividans* et détermination de l'efficacité de
transformation.

Il est apparu que le cosmide contenant l'insert XbaI était moins
stable que celui contenant le fragment PstI EcoRI. C'est donc ce dernier
qui a été retenu sous le nom de pOS700R.

20 La carte du vecteur pOS 700R est représentée sur la Figure 16.

**Exemple 7: Efficacité de transformation des vecteurs intégratifs
(pOS700I) et réplcatifs**

25 **Possibilités**

Rendre la souche de *S. lividans* résistante au thiostrepton par
intégration du plasmide pTO1 portant le marqueur de résistance au
thiostrepton

30 Préparation de protoplastes à partir de *S. lividans* cultivée en
présence de thiostrepton

Avec le vecteur pOS700I, l'efficacité de transformation est
d'environ 3000 transformants par µg d'ADN.

Avec le vecteur pOS700R, l'efficacité de transformation est
d'environ 30 000 transformants par µg d'ADN.

Exemple 8 : Construction d'un vecteur BAC intégratif chez *Streptomyces* et conjugatif

Caractéristiques:

- 5 Réplicatif chez *E. coli*
Transférable par conjugaison de *E. coli* aux *Streptomyces*
Intégratif chez *Streptomyces*
Sélectionnable chez *E. coli* et *Streptomyces*
- 10 Capable d'insérer de grands fragments d'ADN ; il faut souligner qu'il est nécessaire de disposer d'ADN du sol dont la taille est comprise entre 100 et 300kb et non contaminé par des petits fragments. En effet les petits fragments sont très préférentiellement intégrés.
- 15 Doté d'un crible permettant de sélectionner les plasmides portant un insert. Ce crible permet en éliminant les vecteurs refermés sur eux même et non digérés de travailler avec un rapport plus élevé entre vecteur et DNA à insérer ce qui permet d'avoir une meilleure efficacité de clonage pour constituer des banques.

20 **Construction:**

Etape 1 : le vecteur pOSV001

- 25 Clonage d'un fragment PstI-PstI de 800 paires de bases portant l'origine de transfert OriT du réplicon RK2 (Guiney et al., 1983), dans le plasmide pUC19 ouvert par PstI. Cette étape de clonage permet d'obtenir un vecteur transférable de *E. coli* à *Streptomyces* par conjugaison.

La carte du vecteur pOSV 001 est représentée à la Figure 17.

30

Etape 2 : le vecteur pOSV002

Insertion du marqueur Hygromycine (cassette Ω hyg), et sélectionnable chez *Streptomyces*, de sorte que le gène conférant la

résistance à l'hygromycine soit transféré en dernier ce qui permet de s'assurer du transfert complet du BAC avec l'insert d'ADN du sol.

Clonage de la cassette Hygromycine isolée de pHP45 Ω hyg sur un fragment HindIII-HindIII portant le gène de résistance à l'Hygromycine..
Ce fragment est cloné au site PstI (position 201) du vecteur pOSV001.
5 Ce site PstI a été choisi, compte tenu du sens du transfert, pour que le marqueur Hygro soit le dernier transféré lors de la conjugaison. Les extrémités PstI et HindIII sont rendues compatibles après traitement par le fragment Klenow de l'ADN polymérase permettant de générer des
10 "bouts francs". L'orientation du fragment Ω hyg est déterminée en fin de construction.

La carte du vecteur pOSV002 est représentée à la Figure 18.

Etape 3 : le vecteur pOSV010

15

Le fragment XbaI-HindIII isolé du plasmide pOSV002 et contenant le marqueur de résistance à l'hygromycine et l'origine de transfert est cloné dans le plasmide pOSint1 digéré par XbaI et HindIII. L'orientation des sites est telle que le marqueur hygromycine sera
20 toujours transféré en dernier.

Le plasmide pOSint1, représenté à la Figure 8, a été décrit dans l'article de Raynal et al. (Raynal A et al. Mol Microbiol 1998 **28** :333-42).

25 Cette construction permet l'expression de l'intégrase chez *E. coli* et chez *Streptomyces*.

Etape 4 : le vecteur pOSV014

30 Addition d'une "cassette" permettant à terme de sélectionner dans la construction finale les plasmides ayant insérés de l'ADN étranger.

Cette "cassette" porte le gène codant pour le répresseur CI du phage λ et le gène conférant la résistance à la tétracycline. Ce gène porte dans
35 sa région 5' non codante la séquence cible du répresseur. L'insertion

d'ADN dans le site HindIII situé dans la séquence codante de CI conduit à la non production du répresseur et donc à l'expression de la résistance à la tétracycline.

Elle est portée par le plasmide pUN99 décrit dans l'article : Nilsson et al .

5 (Nucleic Acids Res 1983, 11:8019-30)

Un fragment PvuII-HindIII isolé de pOSV010 et contenant les séquences Int, attP, Hygro et oriT est cloné au site MscI de pUN99 .

La carte du vecteur pOSV014 est représentée sur la Figure 19.

10 **Etape 5 : le vecteur pOSV 403, vecteur BAC intégratif et conjugatif**

Cette dernière étape de clonage dans pBAC11 (représenté à la Figure. 20) permet de conférer au plasmide final des caractéristiques de BAC (Bacterial Artificial Chromosome), en particulier l'aptitude à
15 accepter des inserts d'ADN de très grande taille.

Le fragment PstI-PstI du vecteur pOSV014 portant l'ensemble des éléments et fonctions décrits précédemment est cloné dans le plasmide pBAC11 (pBeloBAC11) digéré par NotI. Les extrémités sont rendues compatibles par traitement avec l'enzyme de Klenow.

20 La carte du vecteur pOSV403 est représentée sur la Figure 21. Le schéma de la Figure 21 indique l'orientation retenue.

Etape 6 :

Le vecteur pOSV403 contient les sites HindIII et NsiI. Le site
25 NsiI est assez rare chez *Streptomyces* et présente l'avantage d'être compatible avec PstI. En revanche, le site PstI est fréquent chez *Streptomyces* et peut être utilisé pour effectuer des digestions partielles.

Les clones recombinants portant un insert cloné dans le répresseur CI, et donc inactivant ce répresseur deviennent résistants à la tétracycline. Etant donné que les BACs ne sont présents qu'à raison
30 d'une copie par cellule, il faut sélectionner les clones recombinants avec une dose plus faible de tétracycline que la dose habituelle de 20 µg/ml, par exemple avec une dose de 5 µg/ml. Dans ces conditions il n'y a aucun bruit de fond.

- Il est aussi possible d'utiliser un système développé et commercialisé par la société InVitrogen, dans lequel l'insertion d'ADN dans le vecteur inactive un inhibiteur de la gyrase dont l'expression est toxique pour *E. coli*. Le fragment est préférentiellement isolé à partir du
- 5 vecteur pZErO-2 (<http://www.invitrogen.com/>).

Exemple 9 : Construction d'une banque de *S. alboniger* dans les 2 cosmides intégratif (pOS700I) et réplcatif (pOS700R)

10 **1) - Construction de la Banque**

- Pour évaluer l'efficacité du système de clonage, la voie de biosynthèse de la puromycine de *Streptomyces alboniger*, a été clonée dans les deux cosmides navettes pOS700I et pOS700R. Les gènes de
- 15 la voie de biosynthèse de la puromycine sont portés par un fragment d'ADN BamHI d'environ 15 Kb.

- L'ADN génomique de *Streptomyces alboniger* a été isolé. 90% de cet ADN possède un poids moléculaire compris entre 20 et 150 Kb, déterminé par électrophorèse en champ pulsé.

- 20 Les deux cosmides ont été digérés par l'enzyme BamHI (site unique de clonage).

- Les conditions de digestion partielle BamHI de l'ADN génomique ont été déterminées (50 µg d'ADN et 12 unités d'enzyme, 5 minutes de digestion). Après vérification de la taille par électrophorèse
- 25 en gel d'agarose, l'ADN partiellement digéré a été introduit dans les vecteurs. Dans la ligation, 15 µg d'ADN génomique + 2 µg du vecteur intégratif ou 5 µg du vecteur réplcatif ont été utilisés.

- Chaque mélange de ligation a été utilisé pour l'encapsulation in vitro de l'ADN dans les têtes de bactériophage lambda. Les mélanges
- 30 d'encapsulation (0,5ml) ont été titrés (Vecteur intégratif pOS700I = $7,5 \times 10^5$ cosmides/ml, Vecteur réplcatif = 5×10^4 cosmides/ml).

Les cosmides ont été utilisés pour transfecter *E. coli* et générer ainsi deux banques d'environ 25000 clones résistant à l'ampicilline. L'ADN de l'ensemble de ces clones a été isolé et quantifié.

Pour tester les banques, plusieurs clones ont été choisis, l'ADN purifié et a été digéré par BamHI, afin de vérifier la présence et la taille des inserts. Les clones testés contiennent entre 20 et 35 Kb d'insert de *S. alboniger*.

5

2) - Identification des clones contenant la voie de biosynthèse de la puromycine

Les clones susceptibles de contenir la voie complète de biosynthèse de la puromycine ont été identifiés par hybridation avec une sonde correspondant au gène de résistance à la puromycine, le gène *pac* de 1,1 kb. (Lacalle et al. Gene 1989;79, 375-80)

10

Banque faite dans le Vecteur Intégratif pOS 700I:

Parmi 2000 clones analysés, 9 clones ont hybridé avec la sonde et ils contiennent des inserts d'environ 40 kb.

15

Banque faite dans le Vecteur replicatif pOS 700R:

Parmi 2000 clones analysés, 12 clones ont hybridé avec la sonde; ils contiennent des inserts d'environ 40 kb.

20

En utilisant les données publiées par Tercero et al. (J Biol Chem. 1996; 271, 1579-90), les clones contenant la totalité de la voie de biosynthèse ont été identifiés, après hybridation avec des sondes appropriées. Certains cosmides intégratifs et replicatifs présentent après digestion ClaI-EcoRV un fragment de 12360 paires de bases, ce qui laisse supposer un insert contenant la totalité de la voie de biosynthèse de la puromycine.

25

4) - Vérification de la production de puromycine par les clones résistants (Rhône-Poulenc).

30

a) Matériels et Méthodes

Souches et conditions de culture :

5

Trois clones résistants ont été sélectionnés pour vérifier la production de puromycine. Ils correspondent aux recombinants de *S. lividans* contenant un insert dans le vecteur intégratif pOS700I (G 20) ou un insert dans le vecteur réplicatif (G21 et G22).

10

Des souches de référence ont été utilisées pour s'assurer que les milieux de culture utilisés permettaient cette production. Il s'agit de la souche sauvage *S. alboniger* ATCC 12461, productrice de puromycine et de la souche recombinante *S. lividans* contenant le cluster complet de la puromycine cloné dans le plasmide pRCP11 (Lacalle et al, 1992, the EMBO journal, 11, 785-792) (G23).

Les souches sont ensemencées dans un milieu de culture dont la composition est la suivante :

20	Peptone bactériologique Organotechnie	5 g/l de milieu final
	Extrait de levure Springer	5
	Extrait de viande Liebig	5
	Glucose Prolabo	15
	CaCO3 (1) Prolabo	3
25	NaCl Prolabo	5
	Agar (2) Difco	1

- (1) Les 3g de carbonate sont mélangés à 200ml d'eau distillée puis stérilisés à part. L'addition se faisant après stérilisation.
- 30 (2) L'agar est préalablement fondu dans 100ml d'eau distillée avant d'être ajouté aux autres ingrédients du milieu

pH ajusté à 7,2 avant stérilisation
stérilisation 25 minutes à 121°C

35

- 50 µg/l d'hygromycine et 5 µg/l de thiostrepton sont ajoutés au milieu après stérilisation de façon à maintenir une pression de sélection des clones contenant un insert grâce au gène marqueur présent sur le vecteur (le gène de résistance au thiostrepton étant porté par le plasmide pRCP11).

- 50 ml de milieu de culture liquide, répartis en erlenmeyers de 250 ml, sont ensemencés avec 2 ml de suspension aqueuse de spores et de mycelium de chacune des souches. Les cultures sont incubées pendant 4 jours à 28°C avec une agitation de 220 trs/mn. 50 ml de milieux de production, répartis en erlenmeyers de 250 ml, sont ensuite ensemencés avec 2 ml de ces pré-cultures. Le milieu de production utilisé est un milieu industriel optimisé pour la production de pristinamycine (milieu RPR 201). Les cultures sont incubées à 28°C, avec une agitation de 220trs/mn. Après différents temps d'incubation, un erlenmeyer de chaque culture est amené à pH 11 puis extrait par 2 fois 1 volume de dichlorométhane. La phase organique est concentrée à sec sous pression réduite, puis l'extrait est repris par 10 µl de méthanol. 100 µl de la solution méthanolique sont analysés en CLHP munie d'un détecteur à barrette de diodes dans un système gradient eau-acétonitrile 0,05% TFA V/V sur colonne C18 pour la détection de la puromycine.

b) Résultats

- Les analyses HPLC comparatives à partir des cultures des différentes souches montrent la production de puromycine dans la culture de la souche sauvage à partir de 24 h d'incubation. Une production, bien que plus faible, est aussi nettement détectée à partir de 48 h dans la culture du clone G20 contenant le cosmide pOS700I (figure 23). La puromycine a également été détecté à l'état de trace dans le clone G23 contenant l'opéron complet codant pour le composé dans le plasmide pRCP11. Néanmoins, aucune production n'a été observée dans les cultures des clones G21 et G22 contenant le cosmide pOS700R. Les résultats sont reportés sur la Figure 23.

c) Conclusions

Les résultats obtenus permettent de démontrer l'efficacité du système de clonage développé dans le cosmide pOS700I pour exprimer chez un hôte hétérologue tel que *S. lividans* une voie de biosynthèse complète sous le contrôle de séquences régulatrices qui lui sont propres. D'autre part, ces données valident également le criblage des banques obtenues sur la base de la résistance des clones à la puromycine puisqu'il a conduit à identifier parmi un petit nombre de clones, un recombinant capable d'exprimer la voie de biosynthèse associée au gène de résistance. L'absence de production de puromycine chez les autres clones peut probablement s'expliquer par le clonage d'une partie seulement de l'opéron contenant le gène de résistance mais dépourvue de certaines séquences de régulation, transduction ou transcription nécessaires à la synthèse du composé.

EXEMPLE 10 : - CLONAGE D'ADN DU SOL DANS DES VECTEURS

1) – Préparation de l'ADN du sol à cloner

Les différents fragments d'ADN doivent être purifiés selon leur destination :

Cosmides

La taille des molécules doit être comprise entre 30 et 40kb. Or, l'ADN extrait du sol est hétérogène en taille et comprend des molécules atteignant 200 ou 300kb. Afin d'homogénéiser les tailles, l'ADN est cassé mécaniquement par passage de la solution à travers une aiguille de 0,4mm de diamètre. Les fragments d'une taille voisine de 30kb ne sont pas affectés par ces passages répétés à travers une aiguille et il n'est donc pas nécessaire de faire une séparation par la taille surtout que l'empaquetage dans les particules élimine automatiquement les inserts courts.

BACs

Préparation de l'ADN

5 L'ADN du sol est séparé par électrophorèse en champ pulsé (type CHEF) dans des conditions telles que les fragments compris entre 100 et 300kb sont concentrés dans une bande d'environ 5mm. Ceci est obtenu en réalisant la migration dans un gel à 0,7% d'agarose normal ou 1% d'agarose à bas point de fusion avec un temps de pulsation de 100 secondes pendant 20 heures et à une température
10 de 10°C.

Récupération de l'ADN

15 Deux méthodes sont utilisées, leur choix dépend de la taille des molécules que l'on veut isoler, soit jusqu'à 150kb soit au dessus.

- Jusqu'à 150kb

20 La porosité d'un gel à 0,7% d'agarose permet la sortie de l'ADN par électroélution à condition d'absence totale de bromure d'ethidium. Cet ADN est ensuite manipulé avec des instruments de pipetage à orifice agrandi et hydrophobe pour éviter la fragmentation mécanique des molécules.

- Entre 100 et 300kb

25 La bande contenant les fragments d'une taille entre 100 et 300kb est découpée. Pour la migration un gel d'agarose à 1% et à bas point de fusion est utilisé. Cette propriété permet de fondre le gel à une température supportable pour l'ADN de 65°C et de le digérer ensuite
30 par l'agarase (Agarase commercialisée par la société Boehringer) à une température de 45°C suivant les prescriptions du fournisseur.

2) - Utilisation des cosmides intégratifs pOS700I et replicatifs pOS700R

5 Construction par queues polyA polyT

Principe

Un vecteur cosmide, ouvert à un site de clonage quelconque, est modifié aux extrémités 3' en ajoutant un polynucléotide monotone. D'autre part,
10 l'ADN à cloner est modifié aux extrémités 3' en ajoutant un polynucléotide monotone pouvant s'apparier au précédent.

L'association vecteur-fragment à cloner se fait par ces polynucléotides et la séquence cos du vecteur permet l'empaquetage *in vitro* de l'ADN dans
15 des capsides de phage Lamda.

Préparation du vecteur

Le vecteur utilisé est un vecteur autorépliquatif chez *E. coli* et intégratif
20 chez *Streptomyces*.

Pour *E. coli*, la sélection se fait sur la résistance à l'ampicilline et pour *Streptomyces*, elle se fait sur la résistance à l'hygromycine .

Le cosmide est ouvert à l'un des 2 sites possibles (BamHI ou HindIII) et
25 les extrémités 3' sont rallongées par du polyA avec de la terminale transférase dans les conditions où le fournisseur de l'enzyme prévoit l'addition de 50 à 100 nucléotides.

Préparation de l'ADN à insérer.

30 Les extrémités 3' de l'ADN sont rallongées par du polyT avec de la terminale transférase dans les conditions fournissant un allongement comparable à celui du vecteur. Dans les conditions expérimentales décrites par le fabricant les queues polyA polyT sont longues de 30 à 70
35 bases

Assemblage des molécules et encapsidation in vitro.

- 5 Pour l'assemblage des molécules, on mélange une molécule de vecteur pour une molécule d'ADN inséré. La concentration de l'ADN en masse est de $500 \mu\text{g.ml}^{-1}$.

- Le mélange est encapsidé et l'efficacité de transfection dépend de la souche utilisée comme réceptrice et de l'ADN inséré : nulle avec l'ADN test et la souche DH5 α , l'efficacité est comparable pour les souches 10 SURE et DH10B ; à l'extraction le rendement en ADN est cependant plus élevé avec la souche DH10B.

Construction par déphosphorylation

- 15 L'ADN du sol est mis en bouts francs par élimination des séquences 3' sortantes et remplissage des séquences 5' sortantes. Cette opération est faite avec : enzyme de Klenow, T4 polymérase, les 4 nucléotides triphosphates. Le vecteur cosmétique est digéré par BamHI, puis traité par l'enzyme de Klenow pour le rendre bout franc puis déphosphorylé 20 pour éviter qu'il ne se referme sur lui même. Après ligation, le mélange est encapsidé et transfecté comme précédemment décrit.

3) - Utilisation des pBAC

Principe.

- 25 Le plasmide pBAC conjugatif et intégratif possède les sites HindIII et NsiI comme sites de clonage. L'insertion d'une séquence d'ADN à ces sites inactive le répresseur CI du phage Lambda qui contrôle l'expression du gène de la résistance à la tétracycline. L'inactivation du répresseur rend 30 donc la cellule résistante à cet antibiotique ($5\mu\text{g.ml}^{-1}$). Le clonage à ces sites est facilité par la modification du vecteur et la préparation de l'ADN à cloner.

Préparation du vecteur. Exemple HindIII

Pour que le vecteur ne se referme pas sur lui-même, le site Hind III est
5 modifié : la première base (A) est remise en place pour former une
séquence 5' sortante, qui ne peut pas s'apparier avec ses semblables.
L'opération est effectuée par l'enzyme de Klenow en présence de dATP.

Le succès de l'opération est vérifié en effectuant une ligation du vecteur
10 sur lui-même avant et après traitement à l'enzyme de Klenow. A quantité
d'ADN testé identique, on obtient 3000 clones avant traitement et 60
après traitement.

Préparation de l'ADN (taille comprise entre 100 et 300kb).

15 Mise en bouts francs de l'ADN.

L'ADN est mis en bouts francs par élimination des séquences 3'
sortantes et remplissage des séquences 5' sortantes. Cette
opération est faite avec : enzyme de Klenow, T4 polymérase, les 4
20 nucléotides triphosphates.

Préparation des extrémités. Exemple HindIII

L'addition de l'ADN sur le vecteur se fait au moyen d'oligo-
nucléotides reconnaissant la séquence HindIII modifiée du vecteur.
25 Ils contiennent des sites de restriction rares pour permettre les
clonages ultérieurs (Swal ; NotI). cette technique est dérivée de celle
de : Elledge SJ, Mulligan JT, Ramer SW, Spottswood M, Davis RW.
Proc Natl Acad Sci U S A 1991 Mar 1;88(5):1731-5

Deux oligonucléotides complémentaires sont utilisés :
30 Oligo 1 : 5'-GCTTATTTAAATATTAATGCGGCCGCCCGGG-3'
(SEQ ID N°25)

Oligo 2 : 5'-CCCGGGCGGCCGCATTAATATTTAAATA-3' (SEQ ID
N°26)

Ils sont phosphorylés en 5' par la polynucléotide kinase de T4 en présence d'ATP, après leur hybridation. Cette étape de phosphorylation peut être éliminée en utilisant les oligo-nucléotides déjà phosphorylés.

- 5 La ligation de cet adaptateur double brin avec l'ADN à insérer dans un vecteur est faite par la ligase de T4 en présence d'un très grand excès d'adaptateur (1000 molécules d'adaptateur pour une molécule d'ADN à insérer), en 15 heures à 14°C.

- 10 L'excès d'adaptateur est éliminé par électrophorèse sur un gel d'agarose et les molécules d'intérêt sont récupérées du gel par hydrolyse de celui-ci par de l'agarase ou par électroélution.

Ligation vecteur- ADN .

- 15 La ligation se fait à 14°C sur 15 heures avec 10 molécules de vecteur pour une molécule d'insert.

Transformation.

- 20 La souche réceptrice est la souche DH10B. La transformation se fait par électroporation. Pour exprimer la résistance à la tétracycline, les transformants sont incubés à 37 °C pendant 1 heure en milieu sans antibiotique. La sélection des clones se fait par culture pendant une nuit , sur milieu gélosé LB additionné de tétracycline à 5µg.ml⁻¹.

Exemple 11 : CONJUGAISON CLONE A CLONE ENTRE E. COLI ET STREPTOMYCES

CONJUGAISON ENTRE E COLI SOUCHE S17.1 CONTENANT PPM803 ET STREPTOMYCES LIVIDANS TK 21

30

Introduction

- 35 Il est possible d'effectuer des conjugaisons entre *E. coli* et *Streptomyces* (Mazodier et al, 1989). L'adaptation de cette méthode en développant une technique dite en goutte où l'on mélange 10 µl d'une culture de *E.*

coli contenant un vecteur recombinant à une goutte de *S. lividans* récepteur consiste à réaliser une transformation de clone à clone en s'assurant qu'à la fin de l'opération toute la banque construite dans *E. coli* est introduite dans *S. lividans*. Une transformation en vac amènerait obligatoirement à une multiplication des clones de *Streptomyces* transformants afin d'être pratiquement sûr que la banque dans *E. coli* est complètement représentée dans *S. lividans*.

De plus cette méthode est facilement automatisable.

10 **Essais préliminaires**

Conjugaison entre *E. coli* souche S17.1 contenant le vecteur pOSV303 et *S. lividans* TK21.

Dans ces conditions, on mélange 6×10^6 cellules de *E. coli* avec 2×10^6 spores pré-germées de *S. lividans* dans un volume final de 20 μ l.

Mise au point de la méthode

Il est connu que l'ADN extrait de certains actinomycètes est modifié et de ce fait ne peut être introduit dans certaines souches de *E. coli* sans qu'il soit restreint. La souche de *E. coli* DH10B qui accepte ces ADN n'est pas capable de transférer à *Streptomyces* un plasmide ne contenant que *oriT*, et il est donc nécessaire d'en construire une. Il faudrait y introduire par intégration dans le chromosome un dérivé de RP4 capable de fournir en trans toutes les fonctions nécessaires pour assurer le transfert des clones recombinants contenant l'origine de transfert *oriT*.

Exemple 12 : Construction d'une banque cosmique dans *E. coli* et *Streptomyces lividans* : Clonage de l'ADN du sol

30

L'objectif est la construction d'une librairie d'ADN de grande taille issue de l'environnement, sans étape préalable de culture des microorganismes, dans le but d'accéder aux gènes métaboliques de bactéries (ou de tout autre organisme) que l'on ne sait pas cultiver dans des conditions standard de laboratoire.

35

La procédure décrite a été utilisée pour générer une banque d'ADN dans *Escherichia coli* utilisant le cosmide navette *E. coli*-S. *lividans* pOS7001 et de l'ADN extrait et purifié de la fraction bactérienne d'un sol. Cette dernière méthode permet d'obtenir de l'ADN d'une grande pureté et d'une taille moyenne de 40 kb. Aussi, afin d'éviter pour le clonage une digestion partielle de l'ADN extrait, a été adoptée une stratégie alternative basée sur l'utilisation de l'enzyme terminale transférase qui permet d'ajouter des queues de polynucléotides aux extrémités 3' de l'ADN et du vecteur.

5 5 µg d'ADN ont été extraits de 60 mg de sol de " la Côte Saint André " selon le protocole décrit à l'exemple 3 et traités avec de la terminale transférase (Pharmacia) pour rallonger les extrémités 3' avec un polynucléotide monotone (poly T) (Exemple 10).

15 Le cosmide intégratif pOS7001 est préparé selon le protocole B1, Orsay. Après une étape classique de purification en présence de phénol/chloroforme, l'ADN et le vecteur sont assemblés en mélangeant une molécule de vecteur et une molécule d'ADN inséré. Le mélange est ensuite encapsidé dans les têtes de bactériophages lambda (kit 20 Amersham) qui servent à transfecter *E. coli* DH10B. Les cellules transfectées sont ensuiteensemencées sur milieu LB agar en présence d'ampicilline pour sélection des recombinants résistants à cet antibiotique.

25 Une banque d'environ 5000 clones d'*E. coli* résistants à l'ampicilline a été obtenue. Chaque clone a été ensemencé en milieu LB ou TB + ampicilline dans un puits de microplaque (96 puits) et conservé à -80°C .

30 La séquence aux sites d'insertions des fragments du sol dans le vecteur, pOS7001, générés pendant la construction de la banque a été analysée. Pour cela 17 cosmides de la banques ont été purifiés et séquencés avec une amorce, seq.5' CCGCGAATTCTCATGTTTGACCG 3', qui hybride entre les site BamHI et le site de clonage HindIII présente dans le 35 vecteur.

Les séquences obtenues ont permis d'estimer que la longueur des queues homopolymériques aux points de jonctions est très variable, entre 13 et 60 poly-dA/dT. Au-delà des queues, les séquences des fragments du sol ainsi générées possèdent un pourcentage en G+C entre 53 et 70 %. Des pourcentages si élevés étaient inattendus, mais des résultats similaires ont été déjà reportés sur des préparation brut d'ADN à partir de sol (Chatzinotas A. *et al.*, 1998).

Une stratégie de "pooling" de 48 ou 96 clones à été utilisée pour l'analyse de la richesse microbienne et métabolique. L'ADN cosmique extrait à partir de ces "pools" de clones à été utilisé ensuite pour réaliser des expériences de PCR ou d'hybridation.

15

Exemple 13 : Diversité de l'ADN ribosomique 16S au sein de l'ADN cloné.

a) Matériels et méthodes

Les cosmides de la banque sont extraits à partir de pools de clones par lyse alcaline puis sont purifiés sur gradient de chlorure de césium, afin de prélever la bande d'ADN cosmique sous forme super-enroulée et dans le but d'éliminer tout ADN chromosomique d'*Escherichia coli* pouvant interférer dans l'étude.

Après linéarisation des cosmides par action de la nucléase S1 (50 unités, 30 minutes à 37°C), les séquences d'ADNr 16S contenues dans les pools de clones sont amplifiées dans les conditions standard d'amplification, à partir des amorces universelles 63f (5'-CAGGCCTAACACATGCAAGTC-3') et 1387r (5'-GGCGGWWGTGTACAAGGC-3') définies par MARCHESI *et al.* (1998). Les produits d'amplification d'environ 1.5 kilobases sont purifiés à partir du kit Qiaquick gel extraction (Qiagen) puis directement clonés dans le vecteur pCR II (Invitrogen) chez *Escherichia coli* TOP10, selon les instructions du fabricant. L'insert est alors amplifié à l'aide des amorces M13 Forward et M13 reverse spécifiques au site de clonage du vecteur

PCR II. Les produits d'amplification de taille attendue (environ 1,7 kb) sont analysés par RFLP (Restriction Fragment Length Polymorphism) à l'aide des enzymes CfoI, MspI et BstUI (0,1 unités) afin de sélectionner les clones à séquencer. Les profils de restriction obtenus sont séparés sur gel d'agarose Metaphore 2.5% (FMC Products) contenant 0,4 mg de bromure d'éthidium par ml.

Les séquences d'ADNr 16S sont alors déterminées directement en utilisant les produits PCR purifiés par le kit "Qiaquick gel extraction" à l'aide des amorces de séquençage définies par Normand (1995). Les analyses phylogénétiques sont obtenues en comparant les séquences avec les séquences d'ADNr 16S procaryotes rassemblées dans la base de données Ribosomal Database Project (RDP), version 7,0 MAIDAK et al.(1999) grâce au programme SIMILARITY MATCH, permettant d'obtenir les valeurs de similarité par rapport aux séquences de la base de données.

b) Résultats

Pour déterminer la diversité phylogénétique représentée dans la banque, 47 séquences du gène ARNr16S ont été isolées à partir de pools de 288 clones et ont été séquencées dans leur quasi-totalité. Les résultats sont rapportés dans le tableau 7.

L'analyse des séquences par interrogations des bases de données révèle que la majorité des séquences (>61%) présentent des pourcentages de similarité inférieurs ou égaux à 95% avec des espèces bactériennes identifiées (tableau 7). Sur les 47 séquences analysées, 28 séquences ont pour plus proches voisins des bactéries non cultivées, dont les séquences ont été directement issues d'ADN extrait de l'environnement. La majorité de ces séquences présentent par ailleurs des pourcentages de similarité très faibles (88-95%), 17 séquences sur 28 diffèrent ainsi de plus de 5% par rapport à leurs voisins les plus proches.

Parmi les séquences pouvant être classées dans un groupe phylétique, une majorité de séquences appartiennent à la sous classe a des protéobactéries (18 séquences avec un pourcentage de similarité

compris entre 89 et 99%). Un second groupe de séquences est représenté par la sous classe g des protéobactéries, comprenant 9 séquences dont les pourcentages de similarité varient entre 84 et 99%). Les groupes des b-protéobactéries, d-protéobactéries, firmicutes à bas G+C% et à haut G+C% comprennent respectivement 1, 4, 3 et 5 séquences. Seule une séquence n'a pu être classée au sein des grands groupes taxonomiques bactériens définis : la séquence a22.1(19), son plus proche voisin *Aerothermobacter marianas* (avec une similarité de 89%) étant lui même une souche isolée de l'environnement marin et non classifiée à l'heure actuelle. Enfin, 6 séquences peuvent être classées au sein du groupe des *Acidobacterium/ Holophaga*. Ce groupe présente la particularité de n'être représenté que par deux bactéries cultivées *Acidobacterium capsulatum* et *Holophaga foetida*, l'ensemble de ce groupe étant composé par des bactéries dont seul le gène ARNr16S a été détecté par amplification et clonage à partir d'ADN extrait d'échantillon de l'environnement (principalement de sol), Ludwig *et al* (1997). Les faibles valeurs de similarité entre les différentes séquences composant ce groupe laisse présager une grande hétérogénéité et diversité au sein de ce groupe.

L'ensemble des résultats est représenté sur le tableau 7.

Ces résultats montrent que les séquences contenues dans la banque cosmique proviendraient de micro-organismes non seulement diversifiés phylogénétiquement mais surtout de micro-organismes n'ayant jamais été isolés jusqu'à ce jour.

Les résultats du séquençage des ADN amplifiés ont permis d'établir un arbre phylogénétique des organismes présents dans l'échantillon de sol dont les séquences caractérisées sont originales.

L'arbre phylogénétique représenté à la figure 7 a été réalisé à partir de l'alignement des séquences par le logiciel MASE (Faulner et Jurak, 1988) et corrigé par la méthode des 2 paramètres de Kimura (1980), et à l'aide de l'algorithme Neighbour Joining (Saitou et Nei 1987). L'analyse phylogénétique a permis de comparer les séquences ADNr 16S clonées dans la banque d'ADN du sol, avec les séquences d'ADNr 16S procaryotes rassemblées dans les bases de données Ribosomal

Database Project (RDP), (version 7.0, programme SIMILARITY-MATCH, Maidak *et al* 1999), et dans la base GenBank grâce au logiciel BLAST 2.0 (Atschul *et al*, 1997).

5

Exemple 14 : Présélection génétique de la banque pour l'évaluation de la richesse métabolique

Pour caractériser la banque obtenue en terme de diversité
10 métabolique et identifier les clones contenant des inserts portant des gènes pouvant être impliqués dans des voies de biosynthèse, il a été développé selon l'invention des techniques de criblage génétique basées sur des méthodes PCR afin de détecter et d'identifier des gènes PKS de type I.

15

1 Souches bactériennes, plasmides et conditions de culture

S. coelicolor ATCC101478, *S. ambofaciens* NRRL2420, *S. lactamandurans* ATCC27382, *S. rimosus* ATCC109610, *B. Subtilis*
20 ATCC6633 et *B. licheniformis* THE1856 (collection RPR) ont été utilisés comme sources d'ADN pour les expériences de PCR. *S. lividans* TK24 est la souche hôte utilisée pour le cosmide navette POSI700.

Pour la préparation d'ADN génomique, de suspensions de spores, de protoplastes et pour la transformation de *S. lividans*, on a
25 suivi les protocoles standard décrits dans Hopwood *et al.*(1986). *Escherichia coli* Top10 (INVITROGEN) a été utilisé comme hôte pour le clonage des produits PCR et *E. coli* Sure (STRATAGENE) a été utilisé comme hôte pour le cosmide navette pOS700I. Les conditions de culture de *E. coli*, la préparation de plasmides, la digestion de l'ADN,
30 l'électrophorèse sur gel d'agarose ont été réalisées suivant les procédures standard (Sambroock *et al.*,1996).

2. Amorces PCR:

Les couples d'amorces a1-a2 et b1-b2 ont été définis par
35 l'équipe de N. Bamas-Jacques et leur utilisation a été optimisée pour le

criblage de l'ADN des souches pures et de la banque du sol pour la recherche de gènes codant PKS I)

Tableau 8 :

- 5 **Amorces PCR homologues aux gènes PKS I utilisées pour le criblage de la banque.**

a1 (+)	5' CCSCAGSAGCGCSTSTTCTSGA 3'
a2 (-)	5' GTSCCSGTSCCGTSGSTSTCSA 3'
b1	5' CCSCAGSAGCGCSTCTCTSGA 3'
b2	5' GTSCCSGTSCCGTSGGCCTCSA 3'

10 **Conditions d'amplification :**

- Pour la recherche de PKS I à partir de l'ADN de souches pures, le mélange d'amplification contenait : dans un volume final de 50 µl, entre 50 et 150 ng d'ADN génomique, 200 µM de dNTP, 5mM de MgCl₂ final, 7% de DMSO, tampon 1x Appligene, 0,4 µM de chaque primer et 2,5U de Taq Polymerase Appligène. Les conditions d'amplification utilisées sont : dénaturation à 95°C pendant 2 minutes, hybridation à 65°C pendant 1 minute, élongation à 72°C pendant 1 minute, pour le premier cycle, suivi par 30 cycles où la température est diminuée jusqu'à 58°C comme décrit dans K. Seow *et al.*, 1997. L'étape d'extension finale s'effectue à 72°C pendant 10 minutes.

- Pour la recherche de PKS I à partir de l'ADN de la banque, les conditions PCR sont les mêmes que ci-dessus pour le couple a1-a2 en utilisant entre 100 et 500 ng de cosmide extrait de pools de 48 clones. Pour le couple d'amorces b1-b2, 500ng de cosmides issus de pools de 96 clones ont été utilisés. Le mélange d'amplification contenait 200 µM

de dNTP, 2,5mM de $MgCl_2$ final, 7% de DMSO, tampon 1x Quiagen, 0,4 μM de chaque primer et 2,5U de Taq polymérase Hot-start (Qiagen). Les conditions d'amplification utilisées sont : dénaturation 15' à 95°C suivie par 30 cycles : 1' de dénaturation à 95°C + 1' d'hybridation à 65°C pour le premier cycle et 62°C pour les autres cycles, 1' d'élongation à 72°C, étape d'extension finale de 10' à 72°C.

L'identification des clones positifs à partir des pools de 48 ou 96 clones est effectuée à partir des répliques des microplaques mères correspondantes sur milieu solide ou toute autre méthode standard de réplique.

3 Sous-clonage et séquençage

Les produits PCR des clones identifiés ont été séquencés selon le protocole suivant :

Les fragments sont purifiés sur gel d'agarose (Gel Extraction Kit (Qiagen)) et clonés dans *E.coli* TOP 10 (Invitrogen) à l'aide du kit TOPO TA cloning kit (Invitrogen). L'ADN plasmidique de sous-clones est extrait par lyse alcaline sur un Biorobot (Qiagen) et dialysé durant 2 h sur membrane VS 0,025 μm (Millipore). Les échantillons sont séquencés avec les amorces M13 " Universal " et " Reverse " sur le séquenceur ABI 377 96 (PERKIN ELMER).

4) Résultats

Définition et validation des amorces PCR

Deux régions très conservées de PKS du type I d'actinomycètes, comprenant le site actif de l'enzyme, ont été ciblées pour l'amplification de gènes homologues avec des amorces dégénérées. Ces deux régions correspondent aux séquences PQQR(L)(L)LE et VE(A)HGTGT respectivement.

Des amorces (tableau 8) ont été testées avec l'ADN de souches productrices ou non de macrolides: *Streptomyces coelicolor*,

Streptomyces ambofaciens, producteur de spiramycine, et *Saccharopolyspora erythraea*, producteur de l'érythromycine. Quelles que soient les amorces utilisées, des bandes représentant des fragments d'environ 700 pb et correspondant à la longueur du fragment attendu, ont été obtenues avec toutes les souches.

Ces résultats démontrent la spécificité des amorces a et b pour les gènes PKS I de souches productrices ou de gènes silencieux chez *S. coelicolor*.

Le séquençage des produits PCR obtenus avec le couple d'amorces a1-a2 a permis d'identifier, à partir de la souche *S. ambofaciens*, la séquence d'un gène KS déjà décrite (Demande de brevet européen n° EP0791656) comme appartenant à la voie de biosynthèse du planténolide, précurseur macrolidique de la spiramycine, et deux séquences jamais décrites, Stramb 9 et Stramb12, (voir liste séquences).

En ce qui concerne, *S. erythraea*, la méthode de criblage a permis l'identification d'une séquence de KS (sacery17) identique à celle du KS du module 1 déjà publiée dans Genbank (Numéro d'accès M63677), codant pour la synthétase 1 (DEBS1) du 6-deoxyérythronolide B. Une autre séquence non corrélée à la voie de biosynthèse de l'érythromycine a été identifiée et il s'agit de la séquence SEQ ID N° 32.

Conclusion

Une méthode pour analyser la présence de gènes codant pour les PKS du type I par PCR à partir de différents micro-organismes a été mise au point. La structure très conservée du domaine de la keto-synthétase du type I a permis de réaliser une méthode PCR basée sur l'utilisation d'amorces dégénérées biaisées en GC pour le choix des codons.

Cette approche montre la possibilité d'identifier des gènes ou clusters impliqués dans la voie de biosynthèse des polyketides du type I. Le clonage de ces gènes permet la création d'une collection qui pourra

ensuite être utilisés pour construire des hybrides polyketides. Le même principe peut être appliqué à d'autres classes d'antibiotiques.

Les résultats obtenus ici montrent aussi la présence de gènes pouvant appartenir à des clusters silencieux (SEQ ID N° 30 à 32).

- 5 La présence de clusters silencieux a été déjà documentée dans *S. lividans* et leurs expressions sont déclenchées par des régulateurs spécifiques ou pleiotropiques (Horinouchi et al. ;Umeyama et al. 1996) . Ces résultats suggèrent que la détection de gènes appartenant à des voies dites silencieuses codent en réalité pour des enzymes actives
- 10 capable de diriger, en association avec les autres enzymes spécifiques de la voie, les étapes enzymatiques nécessaires pour la synthèse des métabolites secondaires.

Criblage de la banque

15

Le criblage a été effectué dans les conditions décrites dans la section Matériels et Méthodes en utilisant les couples d'amorces validées à partir de souches productrices.

- En présence du couple d'amorces a1-a2, la taille des produits
- 20 PCR obtenus à partir de l'ADN cosmidique extrait de pools de 48 ou 96 clones était d'environ 700 bp, donc en accord avec les résultats attendus.

L'intensité des bandes obtenues était variable, mais une seule bande d'amplification était présente pour chaque pool d'ADN cible.

- 25 Dans ces conditions, 8 groupes d'ADN cible ont été détectés, correspondant à 9 clones positifs après déréplication.
- Le criblage effectué avec le second couple d'amorces, b1-b2, a donné des résultats d'amplification moins spécifiques puisque de nombreuses bandes satellites étaient observées à côté de la bande de 700 bp.
- 30 Néanmoins, 9 groupes d'ADN cible ont été détectés, correspondant à 14 clones positifs après déréplication à partir de ces clones positifs, l'ADN a été extrait pour les étapes de séquençage et de transformation de *S. lividans*.

Analyse des cosmides

- La digestion des cosmides identifiés par PCR avec l'enzyme DraI, reconnaissant un site riche en AT, libère un fragment supérieur à 23 kb (figure 22). Ceci suggère que la méthode PCR cible préférentiellement l'ADN du sol contenant un haut pourcentage en G+C. Ce résultat est la conséquence de la dégénérescence des amorces utilisées, biaisées en GC pour le choix des codons. Les inserts, comme attendu dans le cas de cosmides, ont une taille supérieure à 23 kb, sauf dans un cas (clone a9B12), ce qui pourrait traduire une certaine instabilité des cosmides. D'autre part, parmi tous les clones sélectionnés, seulement deux d'entre eux, GS.F1 et GS.G11, ont montré le même profil de restriction indiquant un faible taux de redondance dans la banque.
- Les cosmides sélectionnés ont été transférés dans *Streptomyces lividans* par transformation de protoplastes en présence de PEG 1000. L'efficacité de transformation varie entre 30 et 1000 transformants par µg d'ADN cosmidique utilisé.

Séquençage et analyse phylogénétique des gènes PKS I du sol

- La méthode de PCR mise à point sur les souches pures a été utilisée comme décrite sur les cosmides de la banque et 24 clones ont ainsi été identifiés.
- Les produits de PCR d'environ 700 bp obtenus à partir de l'ADN de deux pools (48 clones) et de 8 clones uniques, ont été clonés, après purification sur gel d'agarose, et séquencés. Cela a permis l'identification de 11 séquences.
- L'alignement des séquences protéiques déduites PKSs I du sol avec d'autres PKSs I présentes dans différents micro-organismes (figure 24) montre la présence d'une région très conservée qui correspond à la région consensus du site active de la b-kétoacyl synthétase.

L'analyse des séquences obtenues avec la méthode "Codonpreference" (Gribkov *et al.*, 1984 ; Bibb *et al.*, 1984) a révélé la présence d'un fort biais dans l'usage des codons riches en G+C dans une seule phase de lecture. Les protéines déduites selon cette phase
5 de lecture montrent une forte similarité avec des KSs du type I connus (programme Blast). En particulier, la similarité entre les séquences de KSs du sol et des KSs du cluster de l'érythromycine est d'environ 53%.

Après déréplication d'un pool et identification du clone unique, la séquence du produit PCR obtenu à partir de ce clone est identique à
10 celle du pool ce qui confirme la fiabilité de la méthode utilisée.

L'analyse de la séquence du produit PCR d'un clone a permis l'identification probable de 3 gènes KSI différents. Une de ces séquences (SEQ ID N° 34) a une similarité de 98,7% avec la séquence d'un autre pool, suggérant qu'elles codent pour la même enzyme. Les
15 deux autres séquences sont différentes mais fortement homologues.

Ici, il est décrit pour la première fois le clonage et l'identification dans une banque d'ADN du sol de voies de biosynthèse de métabolites secondaires contenant des gènes codant des KS du type I.

Le pourcentage élevé en G+C des séquences du sol suggère
20 qu'elles puissent dériver de génomes ayant un usage des codons similaire à ceux d'actinomycètes.

Même si les données disponibles dans la littérature sont réduites, on sait que les gènes codant des PKS du type I sont très diversifiés de par leur organisation physique dans le génome, la taille et le nombre de
25 modules contenus dans chaque gène.

La présence de plusieurs domaines provenant d'un seul clone est une confirmation de leur appartenance à des clusters de polyketides asymétriques. Dans un seul cas, deux clones semblent former un contigue puisqu'ils partagent la même séquence pour le domaine KS.

La taille des régions génétiques impliquées dans la synthèse des
30 PKSI varie entre quelques kb pour la pénicilline à environ 120 kb pour la rapamycine. La dimension des inserts cosmidiques peut donc ne pas être suffisante pour l'expression des clusters les plus complexes.

Des gènes codant pour des PKSs I, capables de travailler de
35 façon itérative comme les PKS II et de contrôler la synthèse de

polykétides aromatiques, ont été décrits (Jae-Hyuk et al., 1995). L'étude des clusters des PKSs I du sol pourrait apporter encore des nouveautés dans ce domaine.

5 **5. Identification de 6 gènes codant des polykétides synthases.**

On poursuivant le criblage de la banque de cosmides selon les protocoles décrits dans le présent exemple, les inventeurs ont identifiés
10 un clone de cosmide contenant un insert de 34071 pb contenant plusieurs cadres ouverts de lecture codant pour des polypeptides du type polykétide synthase.

Plus précisément, le cosmide ainsi identifié par criblage de la banque contient six cadres ouverts de lecture codant pour des
15 polypeptides polykétide synthase ou pour des polypeptides fortement apparentés, des peptides synthase non ribosomiques. Une carte détaillée de ce cosmide est représentée à la figure 36.

La séquence nucléotidique complète du cosmide constitue la séquence SEQ ID N°113 du listage de séquences. L'insert d'ADN
20 contenu dans la séquence SEQ ID N°113 constitue la séquence nucléotidique complémentaire (brin -) de la séquence nucléotidique codant pour les différents polykétides synthases.

La séquence nucléotidique de l'insert d'ADN contenue dans le cosmide de la figure 36 qui comprend les cadres de lecture ouverts
25 codant pour les polypeptides polykétides synthases (brin +) est schématisée sur la figure 37 et constitue la séquence SEQ ID N°114 du listage de séquences.

De plus, une carte détaillée des différents cadres de lecture ouverts contenus dans l'insert d'ADN de ce cosmide est représentée à la
30 figure 37.

Les caractéristiques des séquences nucléotidiques comprenant des cadres ouverts de lecture contenus dans l'insert d'ADN de ce cosmide sont détaillées ci-après.

Séquence ORF1

La séquence orf1 comprend un cadre ouvert de lecture partielle d'une longueur de 4615 nucléotides. Cette séquence constitue la
5 séquence SEQ ID N°115, qui débute au nucléotide en position 1 et se termine au nucléotide en position 4615 de la séquence SEQ ID N°114.

La séquence SEQ ID N°115 code pour le polypeptide ORF1 de 1537 acides aminés, ce polypeptide constituant la séquence SEQ ID N°121.

10 Le polypeptide de séquence SEQ ID N°121 est apparenté aux peptides synthases non ribosomiques. Ce polypeptide possède un degré d'identité en acides aminés de 37% avec le peptide synthase de *Anabaena* sp.90 référencé sous le numéro d'accès « emb CACO1604.1 » dans la base de données Genbank.

15

Séquence ORF2

La séquence nucléotidique orf2 a une longueur de 8301 nucléotides et constitue la séquence SEQ ID N°116, qui débute au
20 nucléotide en position 4633 et se termine au nucléotide en position 12933 de la séquence SEQ ID N°114.

La séquence ORF2 code pour le peptide ORF2 d'une longueur de 2766 acides aminés, ce polypeptide constituant la séquence SEQ ID N°122.

25 Le polypeptide de séquence SEQ ID N°122 possède une identité de séquence en acides aminés de 41% avec la séquence MtaD de *Stigmatella aurantiaca* référencée sous le numéro d'accès « gb AAF 19812.1 » de la base de données GENBANK.

Le polypeptide ORF2 constitue une polykétide synthase.

30

Séquence ORF3

La séquence nucléotidique orf3 a une longueur de 5292 nucléotides et constitue la séquence SEQ ID N°117. La séquence SEQ
35 ID N°117 correspond à la séquence qui débute au nucléotide en position

12936 et qui se termine au nucléotide en position 18227 de la séquence SEQ ID N°114.

La séquence nucléotidique SEQ ID N°117 code pour le polypeptide polykétide synthase ORF3 de 1763 acides aminés, ce polypeptide constituant la séquence SEQ ID N°123 selon l'invention.

Le polypeptide ORF3 de séquence SEQ ID N°123 possède une identité de 42% en acides aminés avec la séquence MtaB de *Stigmatella aurantiaca* référencée sous le n° d'accès « gb AAF 19810.1 » de la base de données GENBANK.

Séquence ORF4

La séquence nucléotidique orf4 a une longueur de 6462 nucléotides et constitue la séquence SEQ ID N°118 selon l'invention.

La séquence nucléotidique SEQ ID N°118 correspond à la séquence débutant au nucléotide en position 18224 et se terminant au nucléotide en position 24685 de la séquence nucléotidique SEQ ID N°114.

La séquence nucléotidique SEQ ID N°118 code pour le polypeptide polykétide synthase ORF4 de 2153 acides aminés, ce polypeptide constituant la séquence SEQ ID N°124 selon l'invention.

Le polypeptide ORF4 de séquence SEQ ID N°124 possède une identité de séquence en acides aminés de 46% avec la séquence epoD de *Sorangium cellulosum* référencée sous le n° d'accès « gb AAF62883.1 de la base de données GENBANK.

Séquence ORF5

La séquence nucléotidique orf5 a une longueur de 5088 nucléotides et constitue la séquence SEQ ID N°119 selon l'invention.

La séquence SEQ ID N°119 correspond à la séquence débutant au nucléotide en position 24682 et se terminant au nucléotide en position 29769 de la séquence nucléotidique SEQ ID N°114.

La séquence nucléotidique SEQ ID N°119 code pour le polypeptide polykétide synthase ORF5 de 1695 acides aminés, ce polypeptide constituant la séquence SEQ ID N°125 selon l'invention.

- Le polypeptide polykétide synthase ORF5 de séquence SEQ ID N°125 possède une identité en acides aminés de 43% avec la séquence epod de *Sorangium cellulosum* référencé sous le n° d'accès « gb AAF 62883.1 » de la base de données GENBANK.

Séquence ORF6

- La séquence nucléotidique orf6 a une longueur de 4306 nucléotides, et constitue la séquence SEQ ID N°120 selon l'invention. La séquence nucléotidique SEQ ID N°120 correspond à la séquence débutant au nucléotide en position 29766 et se terminant au nucléotide en position 34071 de la séquence SEQ ID N°114.

La séquence SEQ ID N°120 contient un cadre ouvert de lecture partielle codant pour le polypeptide ORF6 de 1434 acides aminés du type polykétide synthase, ce polypeptide constituant la séquence SEQ ID N°126 selon l'invention.

- Le polypeptide de séquence SEQ ID N°126 possède une identité en acides aminés de 43% avec la séquence epod de *Sorangium cellulosum* référencée sous le numéro d'accès « gb AAF 62883.1 » de la base de données GENBANK.

EXEMPLE 15: Construction de vecteurs navettes de type BAC intégratifs chez *Streptomyces*

Construction de vecteurs navettes du type BAC Intégratifs et conjuguatifs chez *Streptomyces*

15.1 Construction du vecteur pMBD-1

Le vecteur BAC pMBD-1 a été obtenu selon les étapes suivantes:

Etape 1: Le vecteur pOSVO10 a été soumis à une digestion par les enzymes PsTI et BstZ17I afin d'obtenir un fragment nucléotidique de 6,3 kb.

5 **Etape 2:** Le vecteur pDNR-1 a été digéré par les enzymes PstI et PvuII afin d'obtenir un fragment nucléotidique de 4,145 kb.

Etape 3: Le fragment nucléotidique de 6,3 kb provenant du vecteur pOSV017 a été fusionné par ligation au fragment de 4,15 kb
10 provenant du vecteur pDNR-1, afin de produire le vecteur pMBD-1, comme cela est illustré à la figure 30.

15.2 Construction du vecteur pMBD-2

15 Le vecteur pMBD-2 est un vecteur du type BAC contenant une boîte intégrative « ϕ c31 int- Ω hyg ».

ϕ c31 est un phage tempéré à spectre d'hôte large dont le site d'attachement (attP) est bien localisé. Le fragment ϕ c31 int est le fragment minimal de l'actinophage ϕ c31 capable d'induire l'intégration
20 d'un plasmide dans le chromosome de *Streptomyces Lividans*.

Ω hyg est un dérivé de l'interposon Ω capable de conférer la résistance à l'hygromicine chez *E. coli* et *S. Lividans*.

Des vecteurs BAC contenant le système d'intégration ϕ c31 sont décrits par SOSIO et al. (2000) et dans la demande PCT n°99 6734
25 publiée le 29 Décembre 1999.

Le vecteur BAC pMBD-2 a été construit selon les étapes suivantes:

Etape 1: Construction d'une boîte intégrative ϕ c31int Ω hyg dans un plasmide multicopies de *E. coli*.

30 On a tout d'abord amplifié le fragment ϕ c31int à partir du plasmide pOJ436 à l'aide du couple d'amorces suivant:

- L'amorce EV ϕ c31I (SEQ ID N°109) (qui permet d'introduire un site EcoRV à l'extrémité 5' de la séquence ϕ c31) et l'amorce BII ϕ c31F (SEQ ID N°110) (qui permet l'introduction d'un site BglII à l'extrémité 3'
35 de la séquence ϕ c31).

Le fragment Ω hyg a été obtenu par digestion à l'aide de l'enzyme BamHI du plasmide pHP45 Ω hyg décrit par BLONDELET-ROUAULT (1997).

Puis la boîte intégrative ϕ c31 int- Ω hyg a été clonée dans le
5 vecteur pMCS5 digéré par les enzymes BglII et EcoRV.

Etape 2: Construction du vecteur pMBD-2.

Le chromosome artificiel bactérien pBac3.6 décrit par
10 FRENGEN et al. (1999) a été digéré par l'enzyme NheI puis traité par l'enzyme Eco polymérase.

Puis, le vecteur pMCS5 ϕ c31 int- Ω hyg a été digéré par les enzymes SnaBI et EcoRV afin de récupérer la boîte intégrative.

La carte détaillée du vecteur pMBD2 est représentée à la figure
15 31.

15.3 Construction du vecteur pMBD-3.

Le vecteur pMBD-3 est un vecteur intégratif (ϕ c31 int) et
20 conjuguatif (OriT) du type BAC, qui comprend le marqueur de sélection Ω hyg.

La carte du vecteur pMBD-3 ainsi que son procédé de construction sont illustrés à la figure 31.

Le vecteur pMBD-3 a été obtenu en amplifiant le gène OriT à
25 partir du plasmide pOJ436 à l'aide du couple d'amorces de séquences SEQ ID N° 111 et SEQ ID N°112 qui contiennent des sites de restriction pacl.

Le fragment nucléotidique amplifié à l'aide des amorces SEQ ID N°111 et SEQ ID N°112 a été cloné dans le vecteur pMBD2
30 préalablement digéré par l'enzyme Pacl. Le schéma de construction du vecteur pMBD-3 est illustré à la figure 31.

15.4 Construction du vecteur pMBD-4

La carte détaillée du vecteur pMBD-4 est représentée à la figure 32.

- 5 Le vecteur pMBD4 a été obtenu en clonant la boîte intégrative ϕ c31 int- Ω hyg dans le vecteur pCYTAC2.

15.5 Construction du vecteur pMBD-5

- 10 Le schéma de construction du vecteur pMBD-5 est illustré à la figure 33.

- Le vecteur pMBD-5 a été construit par recombinaison du fragment nucléotidique compris entre les deux sites loxP du vecteur pMBD-1 illustré à la figure 33 avec le site loxp contenu dans le vecteur
- 15 BAC désigné pBTP3, une carte détaillée du plasmide pBTP3 étant représentée à la figure 34.

15.6 Construction du vecteur pMBD-6

- 20 Le vecteur pMBD-6 a été construit en recombinant le fragment nucléotidique compris entre les deux sites loxP du vecteur pMBD-1 au niveau du site loxP du vecteur BAC pBeloBac11, comme représenté sur la figure 35.

25

TABEAU 1
Localisation des prélèvements d'échantillon et caractéristiques des sols utilisés
dans les différentes expériences. Les comptes microbiens directs en utilisant
la coloration à l'acridine orange ont été réalisés avant et après broyage du sol

Numéro	Origine	Texture	Quantité (%) de sable Limon Argile	Matière organique (g/kg de sol sec)	pH	Nombre de cellules avant broyage ^a (x10 ⁹ /g poids sol sec)	Nombre de cellules après broyage ^a (x10 ⁹ /g poids sol sec)
1	Australie	Argile sablonneuse	62	49,7	5,8	6,5(0,9)	2,9(1,3)
2	Peyrat le Châ- teau, France	Argile sablonneuse	61	48,2	4,9	7,3(0,6)	5,4(0,8)
3	Côte St-André, France	Terreau sablonneux	50	40,6	5,6	10,0(0,7)	7,5(1,4)
4	Chazay d'Azergue, France	Terreau sablonneux argileux	34	13,9	5,8	7,8(1,1)	4,2(0,6)
5	Guadeloupe, France	Argile	27	17,0	4,8	1,4(0,4)	0,5(0,1)
6	Dombes, France	Terreau sablonneux argileux	20	30,3	4,3	7,5(0,5)	5,6(0,9)

^a n=3; déviation standard entre parenthèses.

TABLEAU 2
Amorces et sondes utilisées pour l'amplification PCR et
l'hybridation sur tâche

Amorce ou sonde	Cible ^{a)}	Séquence (5' à 3')	Référence n°
FGPS431 sonde	Universelle (1382-1406)	ACGGCGGTGTGT(A/G)C	Amann et al., 1995
FGPS122 amorce	Bactéries (6-27)	GGAGAGTTTGTATCATGGCTCAG	Amann et al., 1995
FGPS350 amorce	<i>Streptoporangium</i> (616-635)	CCTGGAGTTAAGCCCCCAAGC	Cette étude
FGPS643 sonde	<i>Streptoporangium</i> (122-142)	GTGAGTAACCTGCCCC(T/C)GACT	Cette étude
R499 amorce	<i>Bacillus anthracis</i>	TTAATTCACTTGCACATGATGGG	Patra et al., 1996
R500 amorce	<i>Bacillus anthracis</i>	AACGATAGCTCCTACATTTGGAG	Patra et al., 1996
C501 sonde	<i>Bacillus anthracis</i>	TTGCTGATACGGTATAGAACCTGGC	Patra et al., 1996
FGPS516 amorce	<i>S. lividans</i> OS48.3	TCCAGATCCTTGACCCGCGAG	Cette étude
FGPS517 amorce	<i>S. lividans</i> OS48.3	CACGACATTGCACTCCACCG	Cette étude
FGPS518 sonde	<i>S. lividans</i> OS48.3	CCGTGAGCCGGATCAG	Cette étude

^{a)} Les positions sur le gène de l'ARNr 16S de *E. coli* sont données entre parenthèses. Pour *B. anthracis* et *S. lividans*, les amorces et les sondes ciblent des séquences chromosomiques spécifiques des organismes respectifs. Ces séquences ne sont pas localisées dans le gène de l'ARNr 16S. La cassette contenant la région cible de *S. lividans* est décrite par Clero-Bardin et al. (non publié).

TABLEAU 3
Quantité d'ADN extrait à partir de différents sols après des traitements
de lyse selon les protocoles n°1 à 5 ($\mu\text{g ADN/g de poids de sol sec} \pm \text{déviati on standard}$)^a
Sols 1, 2, 3 et 6; n = 3; sol 4: n=1.

Sol	Protocole de lyse numéro ^b					
	1	2	3	4a	4b	5a 5b
1. Australie	17+/-2	52+/-2	32+/-5	16+/-3	33+/-2	59+/-1 27+/-0
2. Peyrat	29+/-2	58+/-1	40+/-2	29+/-2	18+/-3	56+/-1 15+/-1
3. Côte St-André	36+/-7	60+/-6	148+/-10	94+/-7	38+/-6	73+/-5 47+/-6
4. Chazay	9	16	ND	32	15	15 70
6. Dombes	4+/-2	26+/-3	43+/-1	61+/-	66+/-1	160+/-7 102+/-5

^a Quantification par imagerie de phosphorescence après hybridation sur tâche avec la sonde universelle FGPS431 (tableau 2).

^b1: Aucun traitement; 2: broyage à sec du sol (G); 3: Cr + homogénéisation Ultraturax (H);

4a: G + H + sonication Microtip (MT); 4b: G + H + sonication Cup Horn (CH); 5a: Cr + H + NT + lyse chimique/enzymatique. Voir aussi figure 1.

^c ND = Non déterminé.

Tableau 4 :
Amorces et sondes utilisées dans la caractérisation
moléculaire des ADN extraits du sol

	Cible (amorce ou sonde)	Séquence (5' - 3')	Position ^a
FGPS 612	Eubactéries (amorce)	C(C/T)AACT(T/C/A)CGTGCCAGCAGCC	506 - 525
FGPS 669	Eubactéries (amorce)	GACGTC(A/G)TCCTCC(C/A/C)CCTTCCTC	1174 - 1194
FGPS 618	Eubactéries (sonde)	ATGG(T/C)TGTCGTCAGCTCG	1056 - 1073
FGPS 614	α -Protéobactéries (sonde)	GTGTAGAGGTGAAATTCGTAG	683 - 703
FGPS 615	β -Protéobactéries (sonde)	CGGTGGATGATGTGGATT	939 - 956
FGPS 616	γ -Protéobactéries (sonde)	AGGTTAAAACTCAAATGA	900 - 917
FGPS 621	Gram plus à bas GC% (sonde)	ATACGTAGGTGGCAAGCG	532 - 549
FGPS 617	Actinomycètes (sonde)	GCCGGGGTCAAACTCGGAGG	1159 - 1149
FGPS 680	Streptomycètes (sonde)	TGAGTCCCCA(A/C/T/C(T/A)CCCCG	1132 - 1149
FGPS 619	Streptosporangium (sonde)	GC TTGGGGCTTAAC TCCAGG	609 - 628

^a : position sur le gène ARNr16S d'*Escherichia coli*

Tableau 5 :
Efficacités d'extraction des cellules bactériennes sur gradient de Nycodenz et quantités d'ADN extrait.
Effet de l'incubation de l'échantillon de sol dans une solution d'extrait de levure 6%,
préalablement à la dispersion et à la centrifugation sur gradient de densité.

	Bactéries extraites		ADN extrait		Actinomycètes ^c cfu /g sol sec	Lyse directe ^d ng ADN/ g sol sec	Lyse en bloc d'agarose ^{ee} ng ADN/ g sol sec
	Microfloire totale ^a bactéries /g sol sec		Microfloire cultivable ^b cfu /g sol sec				
Sans incubation							
Suspension de sol	1.3 10 ⁸ (± 0.1)		6.9 10 ⁸ (± 0.2)		8.6 10 ⁸ (± 1.2)		
Extrait cellulaire	1.9 10 ⁸ (± 0.2)		4.1 10 ⁸ (± 1.5)		2.5 10 ⁸ (± 0.7)	333 (± 35)	221 (± 70)
Efficacité d'extraction	15%		59%		38%		
Avec incubation dans extrait de levure 6%							
Suspension de sol	1.2 10 ⁸ (± 0.1)		7.6 10 ⁷ (± 1.1)		6.6 10 ⁷ (± 0.4)		
Extrait cellulaire	1.6 10 ⁸ (± 0.3)		5.3 10 ⁸ (± 1.4)		3.7 10 ⁸ (± 0.7)	344 (± 30)	341 (± 67)
Efficacité d'extraction	13%		7%		5%		

5

a : Dénombrement microscopique après coloration à l'acridine orange

b : Dénombrement sur milieu solide Trypcase-Soja 10%

c : Dénombrement sur milieu solide HV Agar, après enrichissement 20 minutes à 40°C dans une solution d'extrait de levure 6% - SDS 0,05%.

d : La quantité d'ADN extrait a été évaluée sur gel d'électrophorèse par rapport à une gamme étalon d'ADN de thymus de veau.

e : La quantification a été réalisée après digestion de l'agarose par action d'une b-agarase

10

Tableau 6 :
Caractérisation des ADN extraits en fonction de leur proportion en a, b, et g
sous classes des Protéobactéries, en Gram + à bas GC% et en Actinomycètes ; le signal d'hybridation avec la
sonde procaryote servant de référence 100%.

	a- Protéobactérie s	b- Protéobactérie s	g- Protéobactérie s	Gram+ bas GC%.	Actinomycètes	Streptomycètes
Extraction directe ^a	7.7 % (± 1.4)	5.3 % (± 0.5)	3.3 % (± 0.9)	3.1 % (± 1.7)	14.7 % (± 0.6)	0.8 % (± 0.1)
Extraction indirecte						
Lyse + CsCl	10.9 % (± 1.4)	6.4 % (± 1.4)	14.3 % (± 1.4)	7.9 % (± 1.4)	8.5 % (± 1.4)	3.0 % (± 1.4)
Lyse en bloc	2.9 % (± 1.4)	5.4 % (± 1.4)	11.1 % (± 1.4)	8.0 % (± 1.4)	11.3 % (± 1.4)	2.6 % (± 1.4)
Lyse en bloc + incubation YE	6.3 % (± 1.4)	7.5 % (± 1.4)	17.0 % (± 1.4)	18.1 % (± 1.4)	19.4 % (± 1.4)	4.6 % (± 1.4)

a : broyage dans un broyeur à billes de tungstène, à force centrifuge (protocole d'extraction décrit dans article Frostegard et al.)

YE : solution d'extrait de levure à 6%

Tableau 7:
Diversité des séquences d'ADNr16S contenues dans la banque cosmidique

N° pool (clone n°)	Voisin identifié le plus proche	% de similarité	Voisin le plus proche (classification, référence)	% de similarité
<i>α</i> -Proteobactéries				
a24.1 (2)	Azospirillum brasilense	97.7%		
a4-a6-a7 (7)	Azospirillum brasilense	95.4%		
a4-a6-a7 (23)	Azospirillum brasilense	88.9%	Str L-87 (<i>α</i> -proteobactérie) ¹	89.8%
a52-a53-a5 (15)	Azospirillum lipoferrum	97.6%		
a49-a50-a51 (22)	Agrobacterium tumefaciens	95.0%	Clone JN15d (non publié)	95.5%
a49-a50-a51 (11)	Rhizobium sp	99.7%		
a4-a6-a7 (14)	Rhizobium sp	99.7%		
a30-a31-a32 (7)	Bradyrhizobium japonicum	99.4%		
a19-a20-a26 (5)	Bradyrhizobium genosp	93.3%	Clone DA122 (non publié)	95.9%
a37-a38-a39 (6)	Mesorhizobium sp.	98.9 %		
a19-a20-a26 (9)	Bradyrhizobium sp	90.2%	CloneS-26(<i>α</i> Proteobactérie) ²	95.9%
a46-a47-a48 (14)	Phyllobacterium rubiacearum	97.6 %		

TABLEAU 7 (suite 1)
Diversité des séquences d'ADNr16S contenues dans
la banque cosmique

N° pool (clone n°)	Voisin identifié le plus proche	% de similarité	Voisin le plus proche (classification, référence)	% de similarité
a49-a50-a51 (1)	Caulobacter henricii	97.0%		
a1-a2-a3 (13)	Caulobacter sp.	96.3%		
a52-a53-a5 (8)	Mesorhizobium mediterraneum	92.1%	Clone DA122 (non publié)	94.8%
a34-a35-a36 (3)	Rhodobium orientis	91.8%	Clone (non publié)	95.1%
a1-a2-a3 (4)	Sphingomonas sp.	94.7%	Clone PAD23 (non publié)	95.1%
a8-a9-a10 (13)	Sphingomonas sp.	94.0%		
γ -Protéobactéries				
a40-a41-a42 (13)	Pseudomonas sp	98.9%	clone G26(g-Protéobactérie) ³	99.7%
a15-a16-a17 (12)	Lysobacter antibioticus	94.4%	clone vadinHA77(g-Protéo) ⁴	93.6%
a15-a16-a17 (5)	Xanthomonas sp	93.4%	clone vadinHA77(g-Protéo) ⁴	94.6%
a19-a20-a26 (13)	Luteimonas mephitis	92.9%	Souche rJ15 (non publié)	93.5%
a46-a47-a48 (6)	Methylobacter whittenburyi	88.3%	soil clone S-43(g-Protéo) ²	88.9 %
a11-a12-a13 (11)	Methylobacter whittenburyi	88.3%	soil clone S-43(g-Protéo) ²	88.9%

TABEAU 7 (suite 2)
Diversité des séquences d'ADNr16S contenues dans
la banque cosmétique

N° pool (clone n°)	Voisin identifié le plus proche	% de similarité	Voisin le plus proche (classification, référence)	% de similarité
a34-a35-a36 (5)	Methylobacterium capsulatus	84.9%	soil clone S-12 (d-Protéo) ²	85.6%
a43-a44-a45 (10)	Legionella birninghamensis	88.9%		
a8-a9-a10 (2)	Lamprocystis roseopersicina	87.5%	Clone 2-100C14 (non publié)	95.1%
β-Protéobactéries				
a27-a28-a29 (5)	Rhodocyclus tenuis	90.2%	Clone OPB37 (b-protéo) ⁵	91%
δ-Protéobactéries				
a8-a9-a10 (18)	Nannocystis exedens	92.0%		
a11-a12-a13 (5)	Geobacter sulfurreducens	91.5%		
a27-a28-a29 (8)	Desulfococcus infernus	88.4%	Clone S-31 (d-Protéo) ²	89.1%
a40-a41-a42 (6)	Desulfivibrio aminophilus	85.3%	Clone S-34 (d-Protéo) ²	86.2%
G+ bas GC%				
a23.1	Kurthia zopfii	97.3%		
a25.1	Kurthia zopfii	97.2%		
a18.1 (22)	Kurthia gipsonii	94.4%	G+ bas GC% non identifié RS19 (non publié)	94.8%

TABLEAU 7 (suite 3)
Diversité des séquences d'ADNr16S contenues dans la banque cosmétique

Actinomycètes			
a33. 1	Cellulomonas sp	99.5%	
a 14.7	Streptosporangium longisporum	99.8%	
a 21.7	Arthrobacter polychromogenes	99.2%	
a8-a9-a10 (7)	Arthrobacter oxydans	98.3%	actinomycète non identifié RSW1 (non publié)
a27-a28-a29 (3)	Arthrobacter oxydans	98.9%	actinomycète non identifié RSW1 (non publié)
Acidobacterium ?			
a43-a44-a45 (4)	Holophaga foetida	87.3%	Clone 32-10 (Acidobacterium phylum) ^e
a27-a28-a29 (12)	Desulfuromonas acetexigens	88.8%	Clone Sva0515 (Acidobacterium phylum) ^e
a37-a38-a39 (12)	Desulfuromonas palmitatis	90.3%	Clone Sva0515 (Acidobacterium phylum) ^e
a37-a38-a39 (14)	Halothermothrix orenii	87.5%	Clone i13-7 (Acidobacterium phylum) ^e
a8-a9-a10 (9)	Pelobacter carbinolicus	86.5%	Clone i13-15 (Acidobacterium phylum) ^e
a34-a35-a36 (10)	Nitrococcus mobilis	90.6%	clone RB43 (Acidobacterium phylum) ^e
Non classifié			
a22.1(19)	Aerothermobacter marianus	89.1%	Eubacterie non identifiée (non publié)

GONZALEZ et al (1996) - ²Zhou et al. (1997) - ³Pederson et al (1996) - ⁴Godon et al (1997) - ⁵Hugenoltz et al (1998)

⁶Ludwig (1997)

TABLEAU 9 : Séquences

Désignation	SEQ ID N°
Sondes et amorces	
FGPS431	1
FGPS122	2
FGPS350	3
FGPS643 (T)	4
FGPS643 (C)	5
R499	6
R500	7
C501	8
FGPS516	9
FGPS517	10
FGPS518	11
FGPS612	12
FGPS669	13
FGPS618	14
FGPS614	15
FGPS615	16
FGPS616	17
FGPS621	18
FGPS617	19
FGPS680	20
FGPS619	21
63f	22
1387r	23
Oligo-1 (Exemple 10)	24
Oligo-2 (Exemple 10)	25
A1	26
A2	27
B1	28
B2	29
Acides nucléiques PKS-I	
Amb9	30
Amb12	31
Ery19	32
A9b12	33
A23G1 1-1	34
A26G1 1-2	35
A26G1-10	36

TABLEAU 9 (suite 1): Séquences

Désignation	SEQ ID N°
A35 E4-16	37
A49F1-32	38
A17d2-3	39
A53F11-13	40
A53F11-14	41
A22A 2-11	42
A36E8-1	43
A52E8-2	44
Séquences d'acides aminés PKS-I	
Amb9	45
Amb12	46
Ery19	47
A9b12	48
A23G1 1-1	49
A26G1 1-2	50
A26G1-10	51
A35 E4-16	52
A49F1-32	53
A17d2-3	54
A53F11-13	55
A53F11-14	56
A22A 2-11	57
A36E8-1	58
A52E8-2	59
Séquences ADNr 16S	
a24.1(2)	60
a4.a6.a7 (7)	61
a52.a53.a5(15)	62
a49.a50.a51(11)	63
a4.a6.a7(14)	64
a30.a31.a32(7)	65
a37.a38.a39(6)	66
a46.a47.a48(14)	67
a49.a50.a51(1)	68
a52.a53.a5(8)	69
a8.a9.a10(13)	70
a1.a2.a3(13)	71
a43.a44.a45(10)	72
a27.a28.a29(5)	73

TABLEAU 9 (suite 2): Séquences

Désignation	SEQ ID N°
a23.1	74
a25.1	75
a18.1(22)	76
a33.1	77
a14.7	78
a21.7	79
a8.a9.a10(7)	80
a8.a9.a10(18)	81
a27.a28.a29(3)	82
a34.a35.a36(5)	83
a22.1(19)	84
a11.a12.a13(5)	85
a19.a20.a26(9)	86
a40.a41.a42(6)	87
a27.a28.a29(8)	88
a27.a28.a29(12)	89
a37.a38.a39(12)	90
a46.a47.a48(6)	91
a11.a12.a13(11)	92
a15.a16.a17(12)	93
a15.a16.a17(5)	94
a19.a20.a26(13)	95
a37.a38.a39(14)	96
a8.a9.a10(9)	97
a19.a20.a26(5)	98
a43.a44.a45(4)	99
a1.a2.a3(4)	100
a4.a6.a7(23)	101
a49.a50.a51(22)	102
a8.a9.a10(2)	103
a34.a35.a36(3)	104
a34.a35.a36(10)	105
a40.a41.a42(13)	106

TABLEAU 9 (suite 3) :**Séquences**

Désignation	SEQ ID N°
Amorces	
cos 1 n (exemple 5)	107
cos 2 n (exemple 5)	108
EV ϕ c 311 (exemple 15)	109
BI ϕ c 31F (exemple 15)	110
Amorce 1 (exemple 15)	111
Amorce 2 (exemple 15)	112
Acides nucléiques PKS-I	
Cosmide a2641 (vecteur + insert brin (-))	113
Cosmide a2641 (insert - brin (+))	114
orf1	115
orf2	116
orf3	117
orf4	118
orf5	119
orf6	120
Séquences acides aminés PKS-I	
ORF1	121
ORF2	122
ORF3	123
ORF4	124
ORF5	125
ORF6	126

REFERENCES

- Amann, R. I., W. Ludwig, and K.-H. Schleifer. 1995. Phylogenetic identification and in situ detection of individual microbial cells without cultivation. *Microbiol. Rev.* **59**:143-169.
- Atschul S.F., Madden T.L., Schäffer A.A., Zhang J., Zhang Z., Miller W., Lipman D.J. (1997) " Gapped BLAST and PSI-BLAST : a nex generation of protein databses search programs " *Nucleic Acid Researchs* Vol 25 : 3389-3404
- Atschul SF et al., 1990, *J. Mol Biol*, **215** : 403-410.
- Bakken, L. R. 1985. Separation and purification of bacteria from soil. *Appl. Environ. Microbiol.* **49**:1482-1487.
- Bibb MJ, Findlay PR, Johnson MW, The relationship between base composition and codon usage in bacterial genes and its use for the simple and reliable identification of protein-coding sequences., *Gene* **30**: 1-3, 157-66, Oct, 1984.
- Biesiekierska-Galguen M. (1997) " Atténuation biologique de contaminants xénobiotiques dans le sol - modèle lindane " Diplôme de DEA National de Toxicologie, Université Claude Bernard Lyon I.
- Blondelet-Rouault MH, Weiser J, Lebrihi A, Branny P, Pernodet JL. Institut de Genetique et Microbiologie, URA CNRS 2225, Université Paris XI, Orsay, France. *Gene* 1997 May **6**;190(2):315-7
- Borchert S et al., 1992, *Microbiology Letters*, **92** : 175-180
- BLONDELET-ROUAULT, 1997, *Gene*, 315-317.

- **Boccard, F., Smokvina, T. Pernodet, J.L. Friedmann, A. & Guerinneau M.** (1989) . The integrated conjugative plasmid pSAM2 of *Streptomyces ambofaciens* is related to temperature bacteriophages. *Embo J* 8,973-80
- **Chatzinotas A., Sandaa R-A., Schönhuber W., Amanna R., Daae F.L., Torsvik V., Zeyer J., Hahn D.** (1998) " Analysis of broad-scale differences in microbial community composition of two pristine forest soils " *Systematic and Applied Microbiology* Vol 21 : 579-587
- **Clegg, C. D., K. Ritz, and B. S. Griffiths.** 1997. Direct extraction of microbial community DNA from humified upland soils. *Lett. Appl. Microbiol.* 25:30-33.
- **Clerc-Bardin, S., J.-L. Pernodet, A. Frostegård, and P. Simonet.** Development of a conditional suicide system for a *Streptomyces lividans* strain and its use to investigate conjugative transfer in soil. Submitted.
- **Elledge SJ, Mulligan JT, Ramer SW, Spottswood M, Davis RW.** Department of Biochemistry, Baylor College of Medicine, Houston, TX 77030. *Proc Natl Acad Sci U S A* 1991 Mar 1;88(5):1731-5
- **Engelen, B., K. Meinken, F. Von Wintzingerode, H. Heuer, H.-P. Malkomes, and H. Backhaus.** 1998. Monitoring impact of a pesticide treatment on bacterial soil communities by metabolic and genetic fingerprinting in addition to conventional testing procedures. *Appl. Environ. Microbiol.* 64:2814-2821.
- **Farrelly, V., F. A. Rainey, and E. Stackebrandt.** 1995. Effect of genome size and *rna* gene copy number on PCR amplification of 16S rRNA genes from a mixture of bacterial species. *Appl. Environ. Microbiol.* 61:2798-2801.
- **Faulkner D.V., Jurka J.** (1988) " Multiple Aligned Sequence Editor (MASE) " *Trends in Biochemical Sciences* Vol 13 : 321-322
- **FRENGEN et al.,** 1999, *Genomics*, 58: 250-258.

- **Frostegård, A., Tunlid, A., and Bååth, E.** 1991. Microbial biomass measured as total lipid phosphate in soils of different organic content. *J. Microbiol. Meth.* **14**:151-163.
- **Giddings, G.** 1998. The release of genetically engineered micro-organisms and viruses into the environment. *New Phytol.* **140**:173-184.
- **Gladek, A., and J. Zakrzewska.** 1984. Genome size of *Streptomyces*. *FEMS Microbiol. Lett.* **24**:73-76.
- **Gribskov M, Devereux J, Burgess RR,** The codon preference plot: graphic analysis of protein coding sequences and prediction of gene expression., *Nucleic Acids Res* 12: 1 Pt 2, 539-49, Jan 11, 1984.
- **Guiney et al.,** 1983, *Proc. Natl. Acad. Sci USA*, (12) : 3595-3598.
- **Gourmelen, A. Blondelet-Rouault, M.H. & Pernodet, J.L.** (1998). Characterization of a glycosyl transferase inactivating macrolides, encoded by *gimA* from *Streptomyces ambofaciens*, *Antimicrob Agents Chemother* **42**, 2612-9.
- **Hayakawa, M., and H. Nonomura.** 1987. Humic acid-vitamin agar, a new medium for the selective isolation of soil actinomycetes. *J. Ferment. Technol.* **65**:501-509.
- **Hayakawa, M., Ishizawa K., and H. Nonomura.** 1988. Distribution of rare actinomycetes in Japanese soils. *J. Ferment. Technol.* **66**:367-373.
- **Hickey, R. J., and H. D. Tresner.** 1952. A cobalt containing medium for sporulation of *Streptomyces* species. *J. Bacteriol.* **64**:891-892.
- **Hintermann, G., R., Cramer, Kieser, T., and R. Hütter.** 1981. Restriction analysis of the *Streptomyces glaucescens* genome by agarose gel electrophoresis. *Arch. Microbiol.* **130**:218-222.

- Holben, W. E., J. K. Jansson, B. K. Chelm, and J. M. Tiedje. 1988. DNA probe method for the detection of specific microorganisms in the soil bacterial community. *Appl. Environ. Microbiol.* 54:703-711.
- Hong Fu et al., 1995, Molecular diversity, 1 : 121-124
- Hopwood DA, Bibb M J, Chater K F, Kieser T., Bruton C.J., Kieser H.M., Lydiate D.J., Smith C.P., Ward J.M. and Scrempf H. 1985. Genetic Manipulation of *Streptomyces*. A Laboratory manual. The John Innes Foundation, Norwich, U.K.
- Hopwood, D. A., M. J. Bibb, K. F. Chater, T. Kieser, C. J. Bruton, H. M. Kieser, D. J. Lydiate, C. P. Smith, J. M. Ward, and H. Schrempf. 1985. Genetic manipulation of streptomyces - a laboratory manual. The John Innes Foundation, Norwich, United Kingdom.
- Hohm B. and Collins J., 1980, *Gene*, 11:291-298.
- Horinouchi S., Malpartida F., Hopwood D. et Beppu T., *Mol. Gen. Genet.* (1989) 215 :355-357.
- Imai R., Nagata Y., Fukuda M., Takagi M., Yano K. (1991) " Molecular cloning of a *Pseudomonas paucimobilis* gene encoding a 17-kilodalton pypeptide that eliminates HCl molecules from ?-Hexachlorocyclohexane " *Journal of Bacteriology* Vol 17 ", No21 : 6811-6819
- Jacobsen, C. S., and O. F. Rasmussen. 1992. Development and application of a new method to extract bacterial DNA from soil based on separation of bacteria from soil with cation-exchange resin. *Appl. Environ. Microbiol.* 58:2458-2462.
- Jae-Hyuk Y.U. and Leonard T.J.,1995. Sterigmetscytin biosynthesis in *Aspergillus nidulans* requires a ... type I polyketide synthase. *J. Bacteriol.* (August) : 4792-4800.

- **Ka, J. O., W. E. Holben, and J. M. Tiedje.** 1994. Analysis of competition in soil among 2,4-dichlorophenoxyacetic acid-degrading bacteria. *Appl. Environ. Microbiol.* **60**:1121-1128.
- **Kah-Tong S et al.,** 1997, *J Bacteriol*, G179(23) : **7360-7368**
- **Kimura M.** (1980) " A simple method for estimating evolutionary rates of base substitutions through comparative studies of nucleotide sequences " *Journal of Molecular Evolution* Vol 16 : 111-120
- **Kuske, C. R., K. L. Banton, D. L. Adorada, P. C. Stark, K. K. Hill, and P. J. Jackson.** 1998. Small-scale DNA sample preparation method for field PCR detection of microbial cells and spores in soil. *Appl. Environ. Microbiol.* **64**:2463-2472.
- **Lacalle RA, Pulido D, Vara J, Zalacain M, Jimenez A.** Centro de Biología Molecular (CSIC-UAM), Universidad Autónoma, Canto Blanco, Madrid, Spain. *Gene* 1989 Jul 15;79(2):375-80
- **Lee, S.-Y., J. Bollinger, D. Bezdicek, and A. Ogram.** 1996. Estimation of the abundance of an uncultured soil bacterial strain by a competitive quantitative PCR method. *Appl. Environ. Microbiol.* **62**:3787-3793.
- **Löff, L. G., J. R. Dana, J. V. McArthur, and L. J. Shimkets.** 1995. Comparison of methods of DNA extraction from stream sediments. *Appl. Environ. Microbiol.* **61**:1141-1143.
- **Liesack, W., and E. Stackebrandt.** 1992. Occurrence of novel groups of the domain *Bacteria* as revealed by analysis of genetic material isolated from an Australian terrestrial environment. *J. Bacteriol.* **174**:5072-5078.
- **Liesack, W., P. H. Janssen, F. A. Rainey, N. L. Ward-Rainey, and E. Stackebrandt.** 1997. Microbial diversity in soil: the need for a combined approach using molecular and cultivation techniques. *In* J. D. Van Elsas, J.

T. Trevors, and E. M. H. Wellington (ed.), Modern soil microbiology, Marcel Dekker, Inc., New York. (p 375-439)

- **Lorentz, M. G., and W. Wackernagel.** 1994. Bacterial gene transfer by natural genetic transformation in the environment. *Microbiol. Reviews* 58:563-602.

- **Maidak B.L., Cole J.R., Parker C.T., Garrity G.M., Larsen N., Li B., Lilburn T.G., McCaughey M.J., Olsen G.J., Overbeek R., Pramanik S., Schmidt T.M., Tiedje J.M., Woese C.R.** (1999) "A new project of the RDP (Ribosomal Database Project) " *Nucleic Acids Research* Vol 27 : 171-173

- **Mazodier P. et al.,** 1989, *J. Bacteriol.*, **171(6)** : 3583-3585.

- **Moré, M. I., J. B. Herrick, M. C. Silva, W. C. Ghiorse, and E. L. Madsen.** 1994. Quantitative cell lysis of indigenous microorganisms and rapid extraction of microbial DNA from sediment. *Appl. Environ. Microbiol.* 60:1572-1580.

- **Murakami T, Holt TG, Thompson CJ,** Unité de Génie Microbiologique, Institut Pasteur, Paris, France. *J. Bacteriol* 1989 Mar;171(3):1459-66

- **Nagata Y., Hatta T., Imai R., Kimbara K., Fukuda M., Yano K., Takagi M.** (1993) " Purification and characterization of 7-Hexachlorocyclohexane (?-HCH)dehydrochlorinase (LinA) from *Pseudomonas paucimobilis* " *Bioscience, Biotechnology and Biochemistry* Vol 57 No 9 : 1582-1583

- **Nalin R., Simonet P., Vogel T.M., Normand P.** (1999) " *Rhodanobacter lindaniclasticus* gen.nov., sp. nov., a lindane-degrading bacterium " *International Journal of Systematic Bacteriology* Vol 49 : 19-23

- **Nesme, X., C. Picard, and P. Simonet.** 1995. Specific DNA sequences for detection of soil bacteria. In J. T. Trevors, and J. D. van Elsas (ed.), *Nucleic acids in the environment, methods and application.* Springer Lab Manual. (p 111-139)

- Nilsson B, Uhlen M, Josephson S, Gatenbeck S, Philipson L. Nucleic Acids Res 1983 Nov 25;11(22):8019-30
- Normand P. et al., 1995, Océanis, 21(1) : 31-56
- Ogram, A. V., M. L. Mathot, J. B. Harsh., J. Boyle, and C. A. Pettigrew, JR. 1994. Effects of DNA polymer length on its adsorption to soils. Appl. Environ. Microbiol. 60:393-396.
- Ogram, A., G. S. Saylor, and T. Barkay. 1987. The extraction and purification of microbial DNA from sediments. J. Microbiol. Methods 7:57-66.
- Olsen, R. A., and Bakken, L. R. 1987. Viability of soil bacteria: optimization of the plate-counting technique. Microb. Ecol. 13:59-74.
- Paget, E., L. Jocteur Monrozier, and P. Simonet. 1992. Adsorption of DNA on clay minerals: protection against DNaseI and influence on gene transfer. FEMS Microbiol. Lett. 97:31-40.
- Patra, G., P. Sylvestre, V. Ramisse, J. Thérasse, and J.-L. Guesdon. 1996. Isolation of a specific chromosomal DNA sequence of *Bacillus anthracis* and its possible use in diagnosis. FEMS Immunol. Medical Microbiology 15:223-231.
- Pernodet J.L. Fish, S. Blondelet-Rouault, M.H. & Cundliffe, E. (1996). The macrolide-lincosamide-streptogramin B resistance phenotypes characterized by using a specifically deleted, antibiotic-sensitive strain of *Streptomyces lividans*. Antimicrob Agents Chemother 40, 581, 5.
- Pernodet J.L. , Gourmelen, A., Blondelet-Rouault, M.H. & Cundliffe, E. (1999). Dispensable ribosomal resistance to spiramycin conferred by *srmA* in the spiramycin producer *Streptomyces ambofaciens*. 145, 2355-64.
- Picard, C., C. Ponsonnet, X. Nesme, and P. Simonet. 1992. Detection and enumeration of bacteria in soil by direct DNA extraction and polymerase chain reaction. Appl. Environ. Microbiol. 58:2717-2722.

- **Preud'homme, J., Belloc, A., Charpentié, Y., and Tarridec, P.** 1965. Un antibiotique formé de deux groupes de composants à synergie d'action : la pristinamycine C. R. Acad. Sci. 260 :1309-1312.
- **Priemé, A., J. I. B. Sitaula, A. K. Klemetsson, and L. R. Bakken.** 1996. Extraction of methane-oxidizing bacteria from soil particles. FEMS Microbiol. Ecol. 21: 59-68.
- **Prosser, J.** 1994. Molecular marker systems for detection of genetically engineered micro-organisms in the environment. Microbiol. 140:5-17.
- **Raynal A, Tuphile K, Gerbaud C, Luther T, Guérineau M, Pernodet JL ;** Laboratoire de Biologie et Génétique Moléculaire, Institut de Génétique et Microbiologie, URA CNRS 2225, Université Paris-Sud, Orsay, France. Mol Microbiol 1998 Apr;28(2):333-42
- **Raynald A. Tuphile, K. Gerbaud, C., Luther, T. Guérineau, M. & PERNODET, J.L. (1998).** Structure of the chromosomal insertion site for pSAM2: functional analysis in Escherichia coli. Mol. Microbiol 28, 333-42.
- **Richard, G. M.** 1974. Modifications of the diphenylamine reaction giving increased sensitivity and simplicity in the estimation of DNA. Analytical Biochem. 57:369-376.
- **Romanowski, G., M. G. Lorentz, and W. Wackernagel.** 1993. Use of polymerase chain reaction and electroporation of Escherichia coli to monitor the persistence of extracellular plasmid DNA introduced into natural soils. Appl. Environ. Microbiol. 59:3438-3446.
- **Saitou N., Nei M. (1987) " The Neighbour-Joining method : a new method for reconstructing phylogentic trees " Molecular and Biological Evolution Vol 2 : 112-118**

- Sambrook J., Fritsch E. F. et Maniatis T. 1996. Molecular cloning : a laboratory manual, 2nd ed. Cold spring Harbor Laboratory Press, Cold Spring Harbor, N.Y.
- Sambrook, J., E. F. Fritsch, and T. Maniatis. 1989. Molecular cloning: a laboratory manual, 2nd ed. Cold Spring Harbor Laboratory, Cold Spring Harbor, N. Y.
- Senoo K., Wada H. (1989) " Isolation and identification of an aerobic ?-HCH-decomposing bacterium from soil " *Soil Science, Plant Nutrition* Vol 35, No 1 : 79-87.
- Sezonov, G., Blanc, V., Bamas-Jacques, N., Friedmann, A. Pernodet, J.L. & Guerineau, M.(1997). Complete conversion of antibiotic precursor to pristinamycin IIA by overexpression of *Streptomyces pristinae* biosynthetic genes. *Nat Biotechnol* 15,349-53.
- Shirling, E. B., and D. Gottlieb. 1966. Methods for characterization of *Streptomyces* species. *Int. J. Syst. Bacteriol.* 16:313-340.
- Shizuga et al., 1992, *Proc. Natl. Acad. Sci USA*, 89 : 8794-8797.
- Siefert, J. L., and G. E. Fox. 1998. Phylogenetic mapping of bacterial morphology. *Microbiology* 144:2803-2808.
- Simonet, P., P. Normand, A. Moiroud, and R. Bardin. 1990. Identification of *Frankia* strains in nodules by hybridization of polymerase chain reaction products with strain-specific oligonucleotide probes. *Arch. Microbiol.* 153:235-240.
- Smalla, K., N. Cresswell, L. Mendonca-Hagler, A. Wolters, and D. J. van Elsas. 1993. Rapid DNA extraction protocol from soil for polymerase chain reaction-mediated amplification. *J. Appl. Bacteriol.* 74:78-85.
- SOSIO M. et al. 2000, *Nature Biotechnology*, vol.18,:343-345.

- **Smit, E., P. Leeflang, and K. Wernars.** 1997. Detection of shifts in microbial community structure and diversity in soil caused by copper contamination using amplified ribosomal DNA restriction analysis. *FEMS Microbiol. Ecol.* **23**:249-261.
- **Smokvina T, Mazodier P, Boccard F, Thompson CJ, Guerineau M.** Laboratoire de Biologie et Genetique Moleculaire, Universite Paris-Sud, Orsay, France. *Gene* 1990 Sep 28;94(1):53-9
- **Smolovina, T., Mazodier, P. Boccard, F. Thompson, C.J. & Guerineau, M. (1990).** Construction of a series of pSAM2-based integrative vectors for use in actinomycetes. *Gene* **94**, 53-9.
- **Stackebrandt, E.** 1988. Phylogenetic relationships vs. phenotypic diversity: how to achieve a phylogenetic classification system of the eubacteria. *Can. J. Microbiol.* **34**:552-556.
- **Staneck, J. L., and G. D. Roberts.** 1974. Simplified approach to identification of aerobic Actinomycetes by thin-layer chromatography. *Appl. Microbiol.* **28**:226-231.
- **Stapleton, R. D., S. Ripp, L. Jimenez, S. Cheol-Koh, J. T. Fleming, I. R. Gregory, and G. S. Sayler.** 1998. Nucleic acid analytical approaches in bioremediation: site assessment and characterization. *J. Microbiol. Methods* **32**:165-178.
- **Steffan, R. J., J. Goksøyr, A. K. Bej, and R. Atlas.** 1988. Recovery of DNA from soils and sediments. *Appl. Environ. Microbiol.* **54**:2908-2915.
- **Tebbe, C. C., and W. Vahjen.** 1993. Interference of humic acids and DNA extracted directly from soil in detection and transformation of recombinant DNA from bacteria and a yeast. *Appl. Environ. Microbiol.* **59**:2657-2665.

- **Tercero JA, Espinosa JC, Lacalle RA, Jimenez A.** Centro de Biología Molecular Severo Ochoa, Consejo Superior de Investigaciones Científicas, Madrid, Spain. *J Biol Chem* 1996 Jan 19;271(3):1579-90
- **Thomas J-C., Berger F., Jacquier M., Bernillon D., Baud-Grasset F., Truffaut N., Normand P., Vogel T.M., Simonet P.** (1996) " Isolation and Characterisation of a novel γ -Hexachlorocyclohexane-degrading bacterium " *Journal of Bacteriology* Vol 178, No20 : 6049-6055
- **Torsvik, V. L.** 1980. Isolation of bacterial DNA from soil. *Soil Biol. Biochem.* 12:15-21.
- **Torsvik, V., R. Sørheim, and J. Goksøyr.** 1996. Total bacterial diversity in soil and sediment communities – a review. *J. Ind. Microbiol.* 17:170-178.
- **Tsai, Y.-L., and B. Olson.** 1991. Rapid method for direct extraction of DNA from soil and sediments. *Appl. Environ. Microbiol.* 57:1070-1074.
- **Umeyama T., Tanabe Y., Aigle B.D. et Horinouchi S.,** FEMS (1996) 144 :177-184.
- **Van Elsas, J. D., G. F. Duarte, A. S. Rosado, and K. Smalla.** 1998. Microbiological and molecular biological methods for monitoring microbial inoculants and their effects in the soil environment. *J. Microbiol. Methods* 32:133-154.
- **Van Elsas, J. D., V. Mäntynen, and A. C. Wolters.** 1997. Soil DNA extraction and assessment of the fate of *Mycobacterium chlorophenolicum* strain PCP-1 in different soils by 16S ribosomal RNA gene sequence based most-probable-number PCR and immunofluorescence. *Biol. Fert. Soils* 24:188-195.
- **Volff JN et al.,** 1996, *Mol. Microbiol.*, 21(5) : 1037-1047.
- **Volossiouk, T., E. J. Robb, and R. N. Nazar.** 1995. Direct DNA extraction for PCR-mediated assays. *Appl. Environ. Microbiol.* 61:3972-3976.

- **Wahl GM, Lewis KA, Ruiz JC, Rothenberg B, Zhao J, Evans GA.** Proc Natl Acad Sci U S A 1987 Apr;84(8):2160-4
 - **Waksman, S. A.** 1961. Williams and Wilkins (ed.) The actinomycetes. Classification, identification and description of genera and species. Vol 2. Baltimore.
 - **Ward, D. M., R. Weller, and M. M. Bateson.** 1990. 16S rRNA sequences reveal numerous uncultured microorganisms in a natural community. Nature 344:63-65.
 - **Widmer, F., R. J. Seidler, and L. S. Watrud.** 1996. Sensitive detection of transgenic plant marker gene persistence in soil microcosms. Mol. Ecol. 5:603-613.
 - **Williams, S.T., R. Locci, A. Beswick, D. I. Kurtböke, V. D. Kuznetsov, F. J. Le Monnier, P. F. Long, K. A. Maycroft, R. A. Palma, B. Petrolini, S. Quaroni, J. I. Todd, and M. West.** 1993. Detection and identification of novel actinomycetes. Res. Microbiol. 144:653-656.
 - **Wilson, I. G.** 1997. Inhibition and facilitation of nucleic acid amplification. Appl. Environ. Microbiol. 63:3741-3751.
 - **Woese, C. R.** 1987. Bacterial evolution. Microbiol. Rev. 51:221-271.
- Yannish-Perron et al., 1985 , Gene, **33(1)** : 103-119.
- **Zaslavsky, B. Y.** 1995. Separation of biomolecules, p. 503-667. *In* Aqueous two-phase partitioning. Boris Y. Zaslavsky (ed.) Physical Chemistry and Bioanalytical Applications, Marcel Dekker, Inc., New York.
 - **Zhou, J., M. A. Bruns, and J. M. Tiedje.** 1996. DNA recovery from soils of diverse composition. Appl. Environ. Microbiol. 62:316-322.

REVENDECATIONS

1. Procédé de préparation d'une collection d'acides nucléiques à partir d'un échantillon de sol contenant des organismes, ledit procédé comprenant la succession d'étapes suivante :

- I (a) Obtention de micro-particules par broyage d'un échantillon de sol préalablement séché ou dessiqué, puis mise en suspension des micro-particules dans un milieu tampon liquide ; et
- (b) extraction des acides nucléiques présents dans les micro-particules ; et
- (c)- passage de la solution contenant les acides nucléiques sur un tamis moléculaire, puis récupération des fractions d'élution enrichies en acides nucléiques et passage des fractions d'élution enrichies en acides nucléiques sur un support de chromatographie d'échange d'anions, puis récupération des fractions d'élution contenant les acides nucléiques purifiés.

2. Procédé de préparation d'une collection d'acides nucléiques à partir d'un échantillon de l'environnement contenant des organismes, ledit procédé comprenant la succession d'étapes suivante :

- II (i) Obtention d'une suspension par dispersion de l'échantillon de l'environnement en milieu liquide puis homogénéisation de la suspension par agitation douce; et
- (ii) séparation des organismes et des autres constituants minéraux et/ou organiques de la suspension homogène obtenue à l'étape (i) par centrifugation sur un gradient de densité ; et
- (iii) lyse des organismes séparés à l'étape (ii) et extraction des acides nucléiques; et
- (iv) purification des acides nucléiques sur un gradient de chlorure de césium.

3. Procédé selon la revendication 1, caractérisé en ce que l'étape I- (a) est suivie d'une étape complémentaire de :

- traitement des micro-particules en suspension dans un milieu tampon liquide par sonication ;

4. Procédé selon la revendication 1, caractérisé en ce que l'étape I-(a) est suivie des étapes complémentaires suivantes :

- traitement des micro-particules en suspension dans un milieu tampon liquide par sonication ;
- incubation de la suspension à 37°C après sonication en présence de lysozyme et d'achromopeptidase;
- addition de SDS
- récupération des acides nucléiques.

5. Procédé selon la revendication 1, caractérisé en ce que l'étape I-(a) est suivie des étapes complémentaires suivantes :

- homogénéisation des micro-particules à l'aide d'une étape de mixage violent (vortex) suivie d'une étape de simple agitation ;
- congélation de la suspension homogène suivie d'une décongélation ;
- traitement par sonication de la suspension après décongélation ;
- incubation de la suspension à 37°C après sonication en présence de lysozyme et d'achromopeptidase;
- addition de SDS;

6. Procédé selon l'une des revendications 1 à 5 caractérisé en ce que les acides nucléiques sont des molécules d'ADN.

7. Procédé de préparation d'une collection de vecteurs recombinants, caractérisé en ce que les acides nucléiques obtenus par le procédé selon l'une des revendications 1 à 6 sont insérés dans un vecteur de clonage et/ou d'expression.

8. Procédé selon la revendication 7, caractérisé en ce que les acides nucléiques sont séparés en fonction de leur taille préalablement à leur insertion dans le vecteur de clonage et/ou d'expression.

9. Procédé selon la revendication 7, caractérisé en ce que la taille moyenne des acides nucléiques est rendue sensiblement uniforme par

rupture physique, préalablement à leur insertion dans le vecteur de clonage et/ou d'expression.

10. Procédé selon la revendication 7, caractérisé en ce que le vecteur de clonage et/ou d'expression est du type plasmide.

11. Procédé selon la revendication 7, caractérisé en ce que le vecteur de clonage et/ou d'expression est du type cosmide.

12. Procédé selon la revendication 11, caractérisé en ce qu'il s'agit d'un cosmide réplcatif chez *E. coli* et intégratif chez *Streptomyces*.

13. Procédé selon la revendication 12, caractérisé en ce qu'il s'agit du cosmide pOS700I.

14. Procédé selon la revendication 11, caractérisé en ce qu'il s'agit d'un cosmide conjugatif et intégratif chez *Streptomyces*.

15. Procédé selon la revendication 14, caractérisé en ce que le cosmide est choisi parmi les cosmides pOSV303, pOSV306 et pOSV307.

16. Procédé selon la revendication 11, caractérisé en ce qu'il s'agit d'un cosmide réplcatif à la fois chez *E. coli* et chez *Streptomyces*.

17. Procédé selon la revendication 16, caractérisé en ce qu'il s'agit du cosmide pOS 700R.

18. Procédé selon la revendication 11, caractérisé en ce qu'il s'agit d'un cosmide réplcatif chez *E. coli* et *Streptomyces* et conjugatif chez *Streptomyces*.

19. Procédé selon la revendication 7, caractérisé en ce que le vecteur de clonage et/ou d'expression est du type BAC.

20. Procédé selon la revendication 19, caractérisé en ce qu'il s'agit d'un vecteur BAC intégratif et conjugatif chez *Streptomyces*.

21. Procédé selon la revendication 20, caractérisé en ce que le vecteur est choisi parmi les vecteurs BAC pOSV403, pMBD-1, pMBD-2, pMBD-3, pMBD-4, pMBD-5 et pMBD-6.

22. Procédé de préparation d'un vecteur recombinant de clonage et/ou d'expression, caractérisé en ce que l'étape d'insertion d'un acide nucléique dans le vecteur de clonage et/ou d'expression comprend les étapes suivantes :

- ouvrir le vecteur de clonage et/ou d'expression à un site de clonage choisi, à l'aide d'une endonucléase de restriction appropriée ;
- ajouter un premier acide nucléique homopolymérique à l'extrémité 3' libre du vecteur ouvert ;
- ajouter un second acide nucléique homopolymérique, de séquence complémentaire au premier acide nucléique homopolymérique, à l'extrémité 3' libre de l'acide nucléique de la collection à insérer dans le vecteur ;
- assembler l'acide nucléique du vecteur et l'acide nucléique de la collection par hybridation du premier et du second acide nucléique homopolymérique de séquences complémentaires l'une de l'autre ;
- refermer le vecteur par ligation.

23. Procédé selon la revendication 22, caractérisé en ce que :

- le premier acide nucléique homopolymérique est de séquence poly(A) ou poly (T) ; et
- le second acide nucléique homopolymérique est de séquence poly(T) ou poly(A).

24. Procédé de préparation d'un vecteur recombinant selon l'une des revendications 22 ou 23, caractérisé en ce que la taille de l'acide nucléique à insérer est d'au moins 100 kilobases, préférentiellement d'au moins 200 kilobases.

25. Procédé de préparation d'un vecteur recombinant selon l'une des revendications 22 à 24, caractérisé en ce que l'acide nucléique à insérer est contenu dans la collection d'acides nucléiques obtenus par le procédé selon l'une des revendications 1 à 6.

26. Procédé de préparation d'un vecteur recombinant de clonage et/ou d'expression, caractérisé en ce que l'étape d'insertion d'un d'acide nucléique dans le vecteur de clonage et/ou d'expression comprend les étapes suivantes :

- création de bouts francs sur les extrémités de l'acide nucléique de la collection par élimination des séquences 3' sortantes et remplissage des séquences 5' sortantes ;
- ouverture du vecteur de clonage et/ou d'expression à un site de clonage choisi, à l'aide d'une endonucléase de restriction appropriée ;
- création de bouts francs aux extrémités de l'acide nucléique du vecteur par élimination des séquences 3' sortantes et remplissage des séquences 5' sortantes, puis déphosphorylation des extrémités 5' ;
- Addition d'adaptateurs oligonucléotidiques complémentaires ;
- insertion de l'acide nucléique de la collection dans le vecteur par ligation.

27. Procédé de préparation d'un vecteur recombinant selon la revendication 26, caractérisé en ce que la taille de l'acide nucléique à insérer est d'au moins 100 kilobases, préférentiellement d'au moins 200 kilobases.

28. Procédé de préparation d'un vecteur recombinant selon l'une des revendications 26 ou 27, caractérisé en ce que l'acide nucléique à insérer est contenu dans la collection d'acides nucléiques obtenus par le procédé selon l'une des revendications 1 à 6.

29. Procédé selon l'une des revendications 22 à 28, caractérisé en ce que les acides nucléiques sont insérés tels quels, sans traitement par une ou plusieurs endonucléases de restriction préalablement à leur insertion dans le vecteur de clonage et/ou d'expression.

30. Collection d'acides nucléiques constituée des acides nucléiques obtenus par le procédé selon l'une des revendications 1 à 6.

31. Acide nucléique caractérisé en ce qu'il est contenu dans la collection d'acides nucléiques selon le revendication 30.

32. Acide nucléique selon le revendication 31, caractérisé en ce qu'il comprend une séquence nucléotidique codant au moins un opéron, ou une partie d'un opéron.

33. Acide nucléique selon la revendication 32, caractérisé en ce que l'opéron code pour la totalité ou une partie d'une voie métabolique.

34. Acide nucléique selon la revendication 33, caractérisé en ce que la voie métabolique est la voie de synthèse des polykétides.

35. Acide nucléique selon la revendication 34, caractérisé en ce qu'il est choisi parmi les polynucéotides comprenant les séquences SEQ ID N° 30 à 44 et SEQ ID N° 115 à 120.

36. Acide nucléique selon la revendication 31, caractérisé en ce qu'il comprend la totalité d'une séquence nucléotidique codant pour un polypeptide

37. Acide nucléique selon l'une des revendications 31 à 36, caractérisé en ce qu'il est d'origine procaryote.

38. Acide nucléique selon la revendication 37, caractérisé en ce qu'il provient d'une bactérie ou d'un virus.

39. Acide nucléique selon l'une des revendications 31 à 33 et 36, caractérisé en ce qu'il est d'origine eucaryote.

40. Acide nucléique selon la revendication 39, caractérisé en ce qu'il provient d'un champignon, d'une levure, d'une plante ou d'un animal.

41. Vecteur recombinant caractérisé en ce qu'il est choisi parmi les vecteurs recombinants suivants :

- a) un vecteur comprenant un acide nucléique selon l'une des revendications 35 à 40;
- b) un vecteur obtenu selon le procédé de l'une des revendications 22 à 25 et 29;
- c) un vecteur obtenu selon le procédé de l'une des revendications 26 à 29.

42. Vecteur caractérisé en ce qu'il s'agit du cosmide pOS 700I.

43. Vecteur caractérisé en ce qu'il s'agit du cosmide pOSV303.

44. Vecteur caractérisé en ce qu'il s'agit du cosmide pOSV306.

45. Vecteur caractérisé en ce qu'il s'agit du cosmide pOSV307.

46. Vecteur caractérisé en ce qu'il s'agit du cosmide pOS 700R.

47. Vecteur caractérisé en ce qu'il s'agit du vecteur BAC pOSV403.

48. Vecteur caractérisé en ce qu'il s'agit du vecteur pMBD-1.

49. Vecteur caractérisé en ce qu'il s'agit du vecteur pMBD-2

50. Vecteur caractérisé en ce qu'il s'agit du vecteur pMBD-3.

51. Vecteur caractérisé en ce qu'il s'agit du vecteur pMBD-4.

52. Vecteur caractérisé en ce qu'il s'agit du vecteur pMBD-5.

53. Vecteur caractérisé en ce qu'il s'agit du vecteur pMBD-6.

54. Collection de vecteurs recombinants tels qu'obtenus selon le procédé de l'une des revendications 7 à 21, 25 et 28.

55. Vecteur recombinant de clonage et/ou d'expression caractérisé en ce qu'il est contenu dans la collection de vecteurs recombinants selon la revendication 54.

56. Cellule hôte recombinante comprenant un acide nucléique selon l'une des revendications 31 à 40 ou un vecteur recombinant selon la revendication 55.

57. Cellule hôte recombinante selon la revendication 56, caractérisée en ce qu'il s'agit d'une cellule procaryote ou eucaryote.

58. Cellule hôte recombinante selon la revendication 57, caractérisée en ce qu'il s'agit d'une bactérie.

59. Cellule hôte recombinante selon la revendication 58, caractérisée en ce qu'il s'agit d'une bactérie choisie parmi *E. coli* et *Streptomyces*.

60. Cellule hôte recombinante selon la revendication 58, caractérisée en ce qu'il s'agit d'une levure ou d'un champignon filamenteux.

61. Collection de cellules hôtes recombinantes, chacune des cellules hôtes constitutives de la collection comprenant un acide nucléique de la collection d'acides nucléiques selon la revendication 30.

62. Collection de cellules hôtes recombinantes, chacune des cellules hôtes constitutives de la collection comprenant un vecteur recombinant selon l'une des revendications 41 ou 55.

63. Procédé de détection d'un acide nucléique de séquence nucléotidique déterminée, ou de séquence nucléotidique structurellement apparentée à une séquence nucléotidique déterminée, dans une collection

de cellules hôtes recombinantes selon l'une des revendications 61 ou 62, caractérisé en ce qu'il comprend les étapes suivantes :

- mettre en contact la collection de cellules hôtes recombinantes avec un couple d'amorces hybridant avec la séquence nucléotidique déterminée ou hybridant avec la séquence nucléotidique structurellement apparentée à une séquence nucléotidique déterminée ;
- réaliser au moins trois cycles d'amplification ;
- détecter l'acide nucléique éventuellement amplifié..

64. Procédé de détection d'un acide nucléique de séquence nucléotidique déterminée, ou de séquence nucléotidique structurellement apparentée à une séquence nucléotidique déterminée, dans une collection de cellules hôtes recombinantes selon l'une des revendications 61 ou 62, caractérisé en ce qu'il comprend les étapes suivantes :

- mettre en contact la collection de cellules hôtes recombinantes avec une sonde hybridant avec la séquence nucléotidique déterminée ou hybridant avec une séquence nucléotidique structurellement apparentée à la séquence nucléotidique déterminée ;
- détecter l'hybride éventuellement formé entre la sonde et les acides nucléiques compris dans les vecteurs de la collection.

65. Procédé pour identifier la production d'un composé d'intérêt par une ou plusieurs cellules hôtes recombinantes dans une collection de cellules hôtes recombinantes selon l'une des revendications 61 ou 62, caractérisé en ce qu'il comprend les étapes suivantes :

- culture des cellules hôtes recombinantes de la collection dans un milieu de culture approprié ;
- détection du composé d'intérêt dans le surnageant de culture ou dans le lysat cellulaire d'une ou plusieurs des cellules hôtes recombinantes cultivées.

66 Procédé pour sélectionner une cellule hôte recombinante produisant un composé d'intérêt dans une collection de cellules hôtes

recombinantes selon l'une des revendications 61 ou 62, caractérisé en ce qu'il comprend les étapes suivantes :

- culture des cellules hôtes recombinantes de la collection dans un milieu de culture approprié ;
- détection du composé d'intérêt dans le surnageant de culture ou dans le lysat cellulaire d'une ou plusieurs des cellules hôtes recombinantes cultivées.
- sélection des cellules hôtes recombinantes produisant le composé d'intérêt.

67. Procédé pour la production d'un composé d'intérêt caractérisé en ce qu'il comprend les étapes suivantes :

- cultiver d'une cellule hôte recombinante sélectionnée selon le procédé de la revendication 66;
- récupérer et, le cas échéant, purifier, le composé produit par ladite cellule hôte recombinante.

68. Composé d'intérêt caractérisé en ce qu'il est obtenu selon le procédé de la revendication 67.

69. Composé selon la revendication 68, caractérisé en ce qu'il s'agit d'un polykétide.

70. Polykétide caractérisé en ce qu'il est produit grâce à l'expression d'au moins une séquence nucléotidique comprenant une séquence choisie parmi les séquences SEQ ID N°30 à 44 et SEQ ID N°115 à 120.

71. Composition comprenant un polykétide selon la revendication 69 ou 70.

72. Composition pharmaceutique comprenant une quantité pharmacologiquement active d'un polykétide selon la revendication 69 ou 70, en association avec un véhicule pharmaceutiquement compatible.

73. Procédé de détermination de la diversité des acides nucléiques contenus dans une collection d'acides nucléiques et tout particulièrement d'une collection d'acides nucléiques provenant d'un échantillon de l'environnement, préférentiellement d'un échantillon du sol, ledit procédé comprenant les étapes suivantes:

- mise en contact des acides nucléiques de la collection d'acides nucléiques à tester avec un couple d'amorces oligonucléotidiques hybridant à toute séquence d'ADN ribosomal 16 S bactérien;
- réalisation d'au moins trois cycles d'amplification;
- détection des acides nucléiques amplifiés à l'aide d'une sonde oligonucléotidique ou d'une pluralité de sondes oligonucléotidiques, chaque sonde hybridant spécifiquement avec une séquence d'ADN ribosomal 16 S commune à un règne, un ordre, une sous-classe ou un genre bactérien;
- le cas échéant, comparer les résultats de l'étape de détection précédente avec les résultats de détection, à l'aide de la sonde ou de la pluralité de sondes, d'acides nucléiques de séquence connue constituant une gamme étalon.

74. Procédé selon la revendication 73, caractérisé en ce que le couple d'amorces hybridant à toute séquence d'ADN ribosomal 16 S bactérien est constitué de l'amorce FGPS 612 (SEQ ID N°12) et de l'amorce FGPS 669 (SEQ ID N°13).

75. Procédé selon la revendication 73, caractérisé en ce que le couple d'amorces hybridant à toute séquence d'ADN ribosomal 16 S bactérien est constitué de l'amorce 63 f (SEQ ID N°22) et de l'amorce 1387 (SEQ ID N°23).

76. Acide nucléique comprenant une séquence nucléotidique d'ADNr 16S choisie parmi les séquences possédant au moins 99% d'identité en nucléotides avec les séquences SEQ ID N° 60 à SEQ ID N° 106.

77. Procédé de production d'une polykétide synthase de type I, ledit procédé de production comprenant les étapes suivantes:

- obtention d'une cellule hôte recombinante comprenant un acide nucléique codant pour une polykétide synthase de type I comprenant une séquence nucléotidique choisie parmi les séquences SEQ ID N°33 à SEQ ID N°44, SEQ ID N°30 à SEQ ID N°32 et SEQ ID N° 115 à SEQ ID N°120.

- culture des cellules hôtes recombinantes dans un milieu de culture approprié;

- récupération et, le cas échéant, purification de la polykétide synthase de type I à partir du sumageant de culture ou du lysat cellulaire.

78. Polykétide synthase comprenant une séquence d'acides aminés choisie parmi les séquences SEQ ID N°45 à 59 et SEQ ID N° 121 à SEQ ID N°126.

79. Anticorps dirigé contre une polykétide synthase selon la revendication 78.

80. Procédé de détection d'une polykétide synthase de type I ou d'un fragment peptidique de cette enzyme, dans un échantillon, ledit procédé comprenant les étapes de:

- a) mettre en contact un anticorps selon la revendication 79 avec l'échantillon à tester;

- b) détecter le complexe antigène/anticorps éventuellement formé.

81. Nécessaire de détection d'une polykétide synthase de type I dans un échantillon comprenant:

- a) un anticorps selon la revendication 79;

- b) le cas échéant, des réactifs nécessaires à la détection du complexe antigène/anticorps éventuellement formé.

Sol sec et tamisé

-
- ```
graph TD; A[Sol sec et tamisé] --> B[1. Détermination de l'ADN extracellulaire:
Homogénéisation de la suspension de sol (Vortex)]; A --> C[2. Broyage du sol sec.]; C --> D[3. Homogénéisation de la suspension de sol]; D --> E[4a. Sonication à l'aide d'une micro-pointe]; D --> F[4b. Sonication à l'aide du dispositif Cup Horn]; E --> G[5a. Lyse par le lysozyme et le SDS]; F --> H[5b. Lyse sur le lysozyme et le SDS];
```
1. Détermination de l'ADN extracellulaire:  
Homogénéisation de la suspension de sol (Vortex)
  2. Broyage du sol sec.
    3. Homogénéisation de la suspension de sol
      - 4a. Sonication à l'aide d'une micro-pointe
        - 5a. Lyse par le lysozyme et le SDS
      - 4b. Sonication à l'aide du dispositif Cup Horn
        - 5b. Lyse sur le lysozyme et le SDS

FIG. 1

2/38

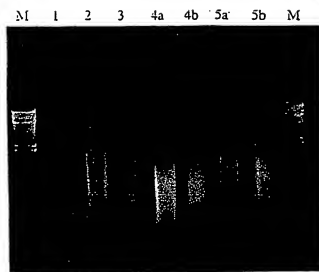


Figure 2



Figure 3

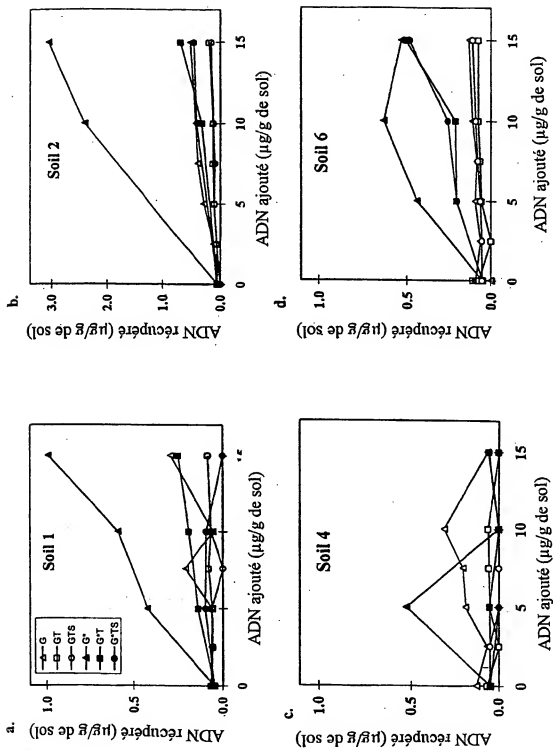


Figure 4



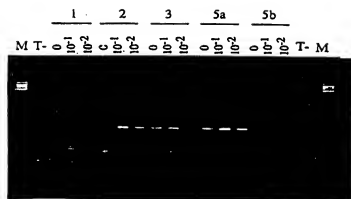


Figure 5

6/38

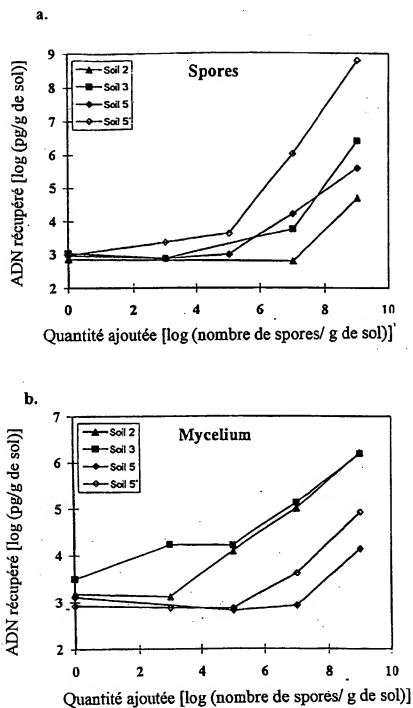


Figure 6

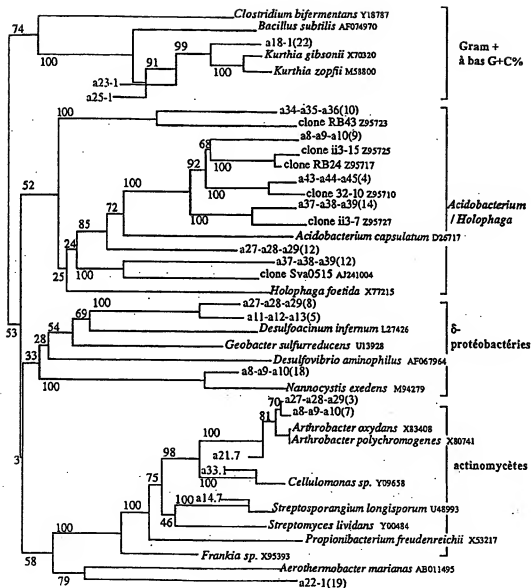


Figure 7- a)

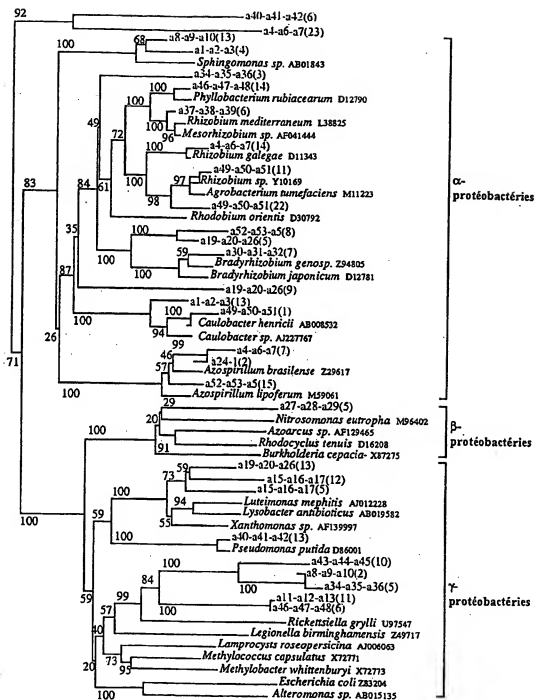


Figure 7 - b)

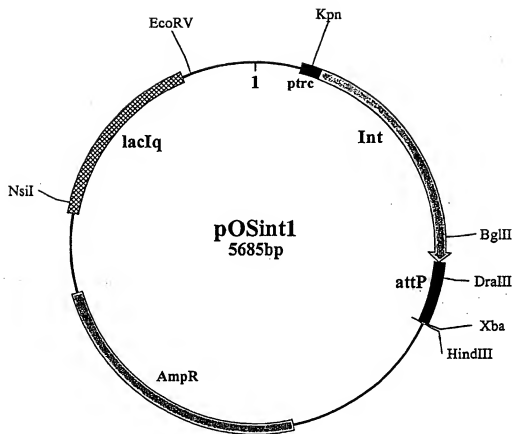


Figure 8

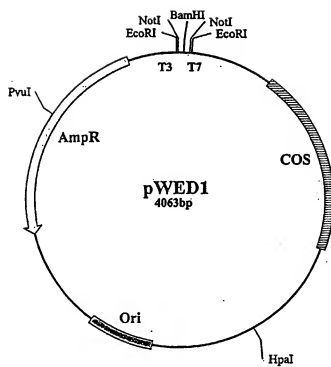


FIGURE 9

11/38

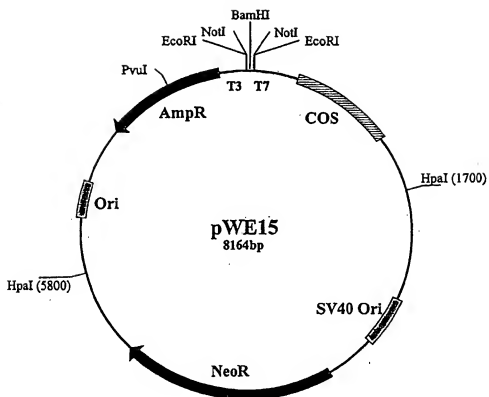


Figure 10

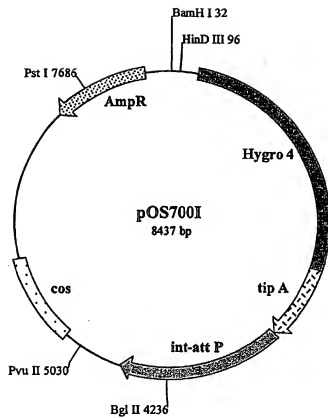


Figure 11



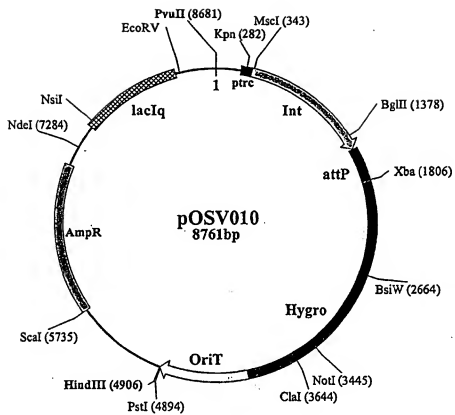
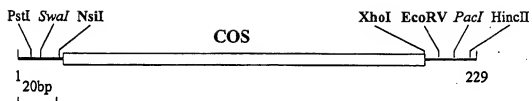


Figure 12

14/38



PCR COS

Figure 13

15/38

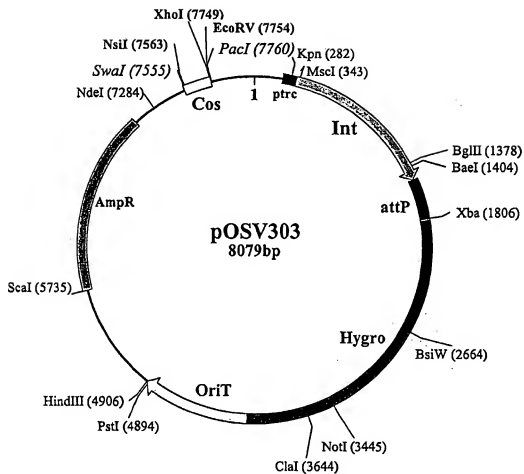


Figure 14

16/38

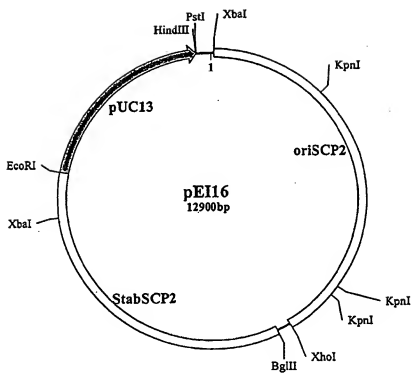


FIGURE 15

17/38

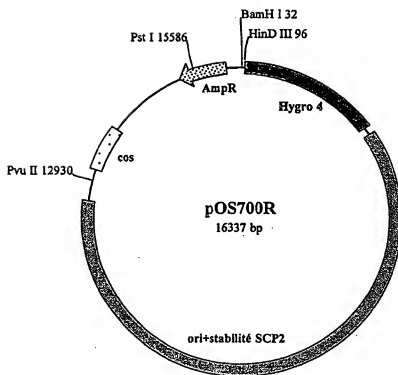


Figure 16

18/38

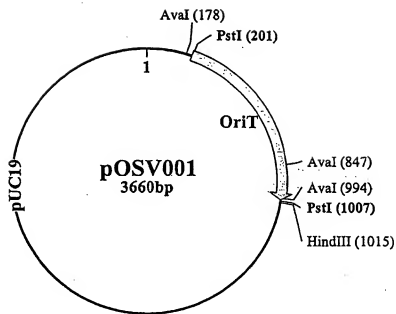


Figure 17

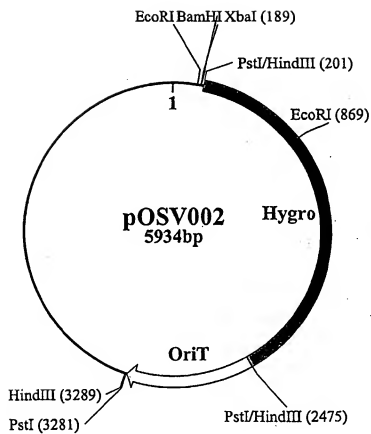


Figure 18

20/38

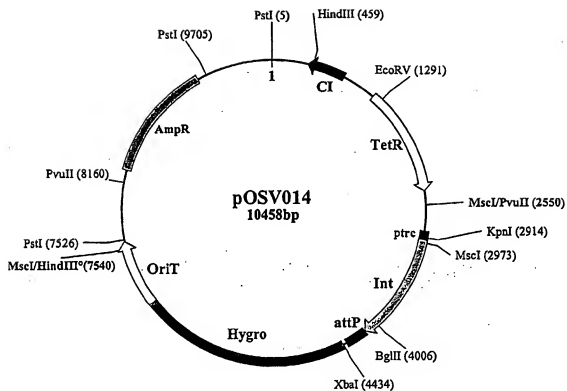


Figure 19



21/38

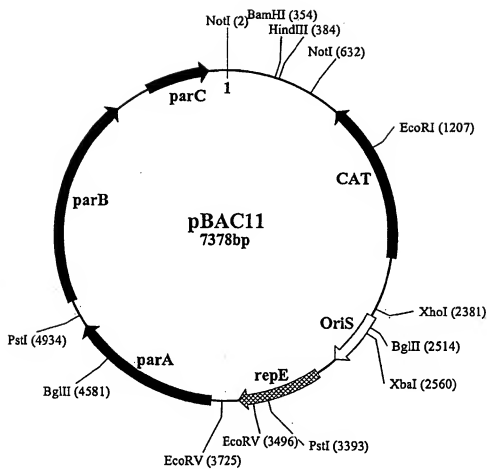


Figure 20

22/38

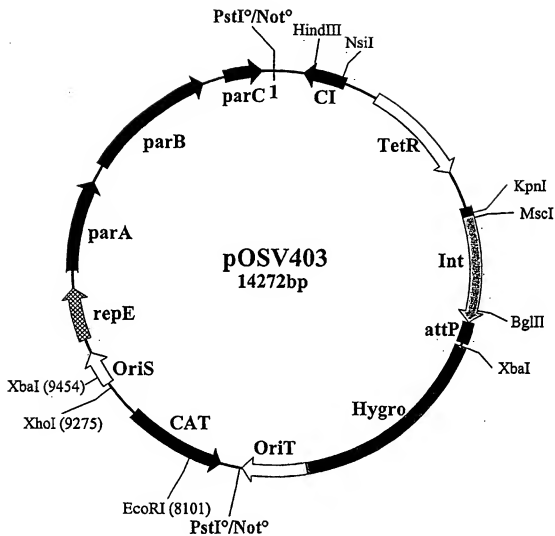


Figure 21

23/38

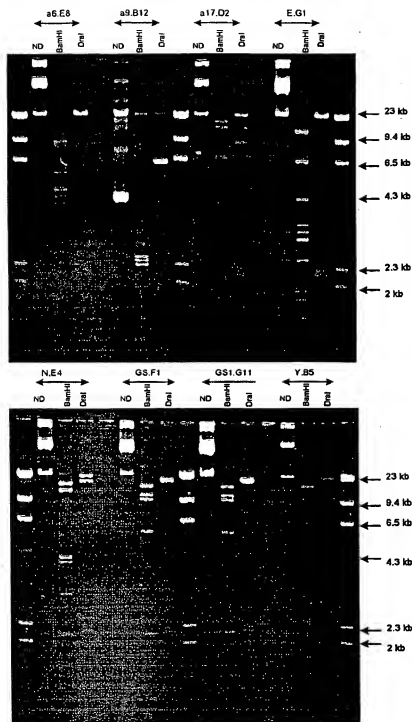


Figure 22

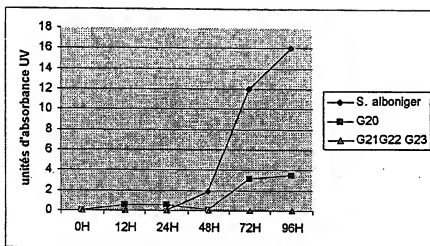


Figure 23

|                        |            |            |                       |
|------------------------|------------|------------|-----------------------|
| 101                    |            |            | 150                   |
| sol_a26G1-1            | KDFLATRVSY | KLNLRGPSLT | VQTACSTSLV SVVMACESLQ |
| sol_a46B5-31           | KDYLPTRVSY | KLNLRGPSLA | VQSACSTGLV AVCQAIQNLQ |
| sol_a9B12-3            | KDFIATRTAY | KLNLRGPAMA | VGACSTSLV AVHEACQALR  |
| sol_a49F1-32           | KDFLATRTAY | KLNLRGPAMT | VQTACSSSLV AVHVAAQSL  |
| <i>B. subtilis</i>     | SGTIPTMISH | KLGRLGPSYF | VHANCSSSLI GLHSAYKSL  |
| stramb12               | GSVLSGRIAY | TFGQLQGPV  | VDACSSSLV ALHLAAQALP  |
| stramb9                | AAVLSGRVSY | AFGLEGPV   | VDACSSSLV ALHLAAQALR  |
| EryA (module 1)        | TSVASGRVSY | TIGLEGPV   | VDACSSSLV AVHLACQSLR  |
| sol_a26G1-2            | FSTAAGRISY | LLGLQGPV   | VDACSSSLV AVHLACQSLR  |
| sol_a53F11-13          | LNAAGRISY  | VLGLQGPSMA | VDACSSSLV AIHLACQSLR  |
| sol_a53F11-14          | HSIAAGRISY | VLGLQGPAMA | VDACSSSLV AIHLACQSLR  |
| sol_a26G1-10           | HSMLANRISY | LLDLRGPSMA | VDACSSALV AVHLACQSLR  |
| sol_a36E8-1            | LSIAANRISY | TFDFRGPSLA | VDACSSSLV AIHLACQSVR  |
| <i>M. tuberculosis</i> | MSILANRISY | FLDLRGPSVA | VDACSSSLV AIHLACQSVR  |
| sacery19               | LSIIPARISY | FLGLRGPDMT | LNTACSSALV AMHQARQSIL |
| sol_a17D2-3            | LAVVANRISY | IYDLRGPSLT | VDACSSSLV ALHQAVEALR  |
| <i>fas</i> humane      | RAMMANRLSP | FFDFRGPSIA | LDTACSSSLM ALQNAVQAIH |

Figure 24

26/38

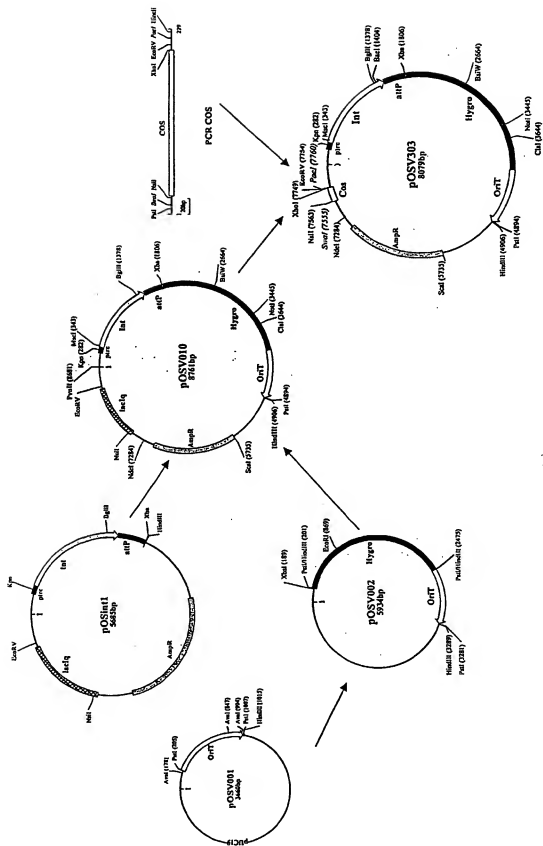
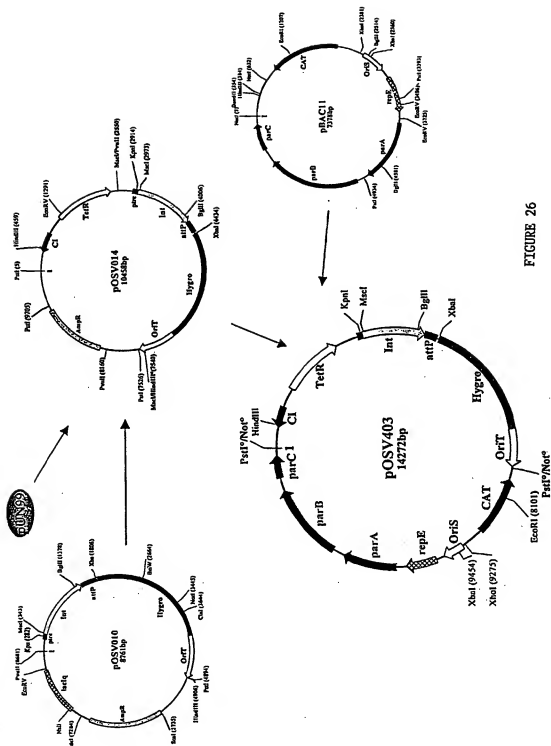


FIGURE 25

27/38



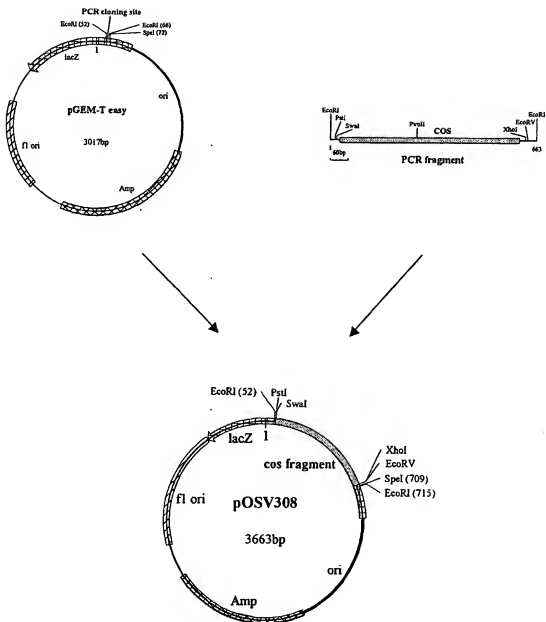


FIGURE 27



29/38

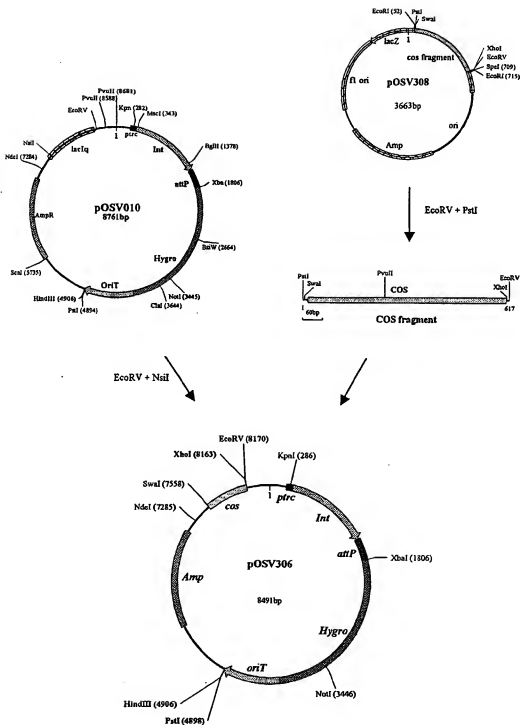


FIGURE 28

30/38

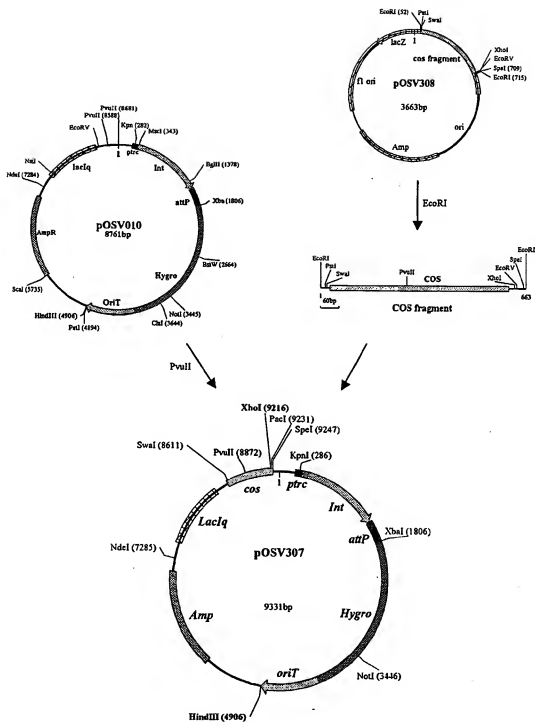


FIGURE 29

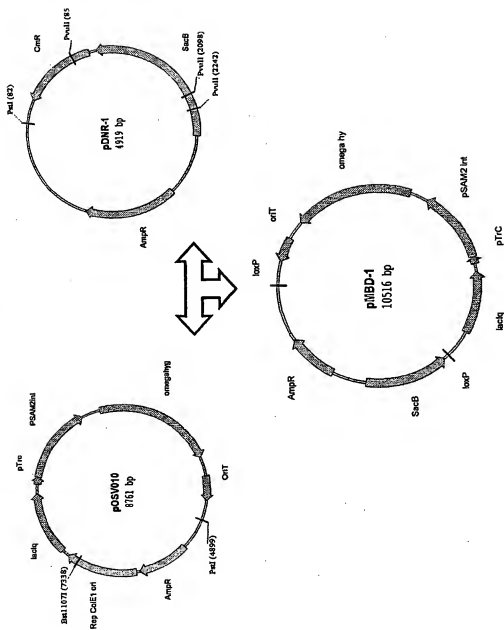


FIGURE 30

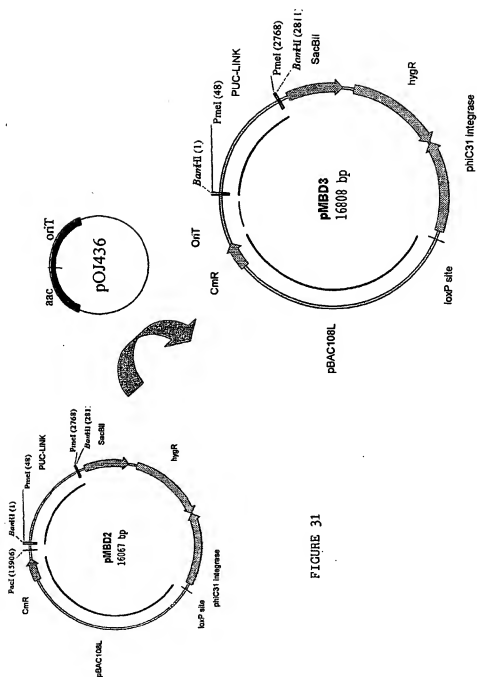


FIGURE 31

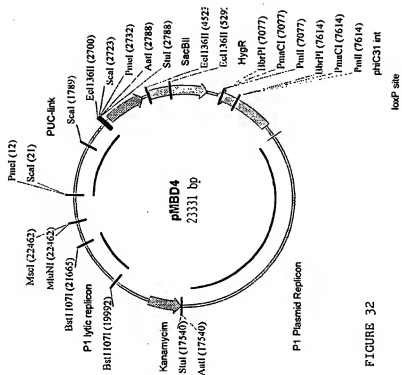


FIGURE 32

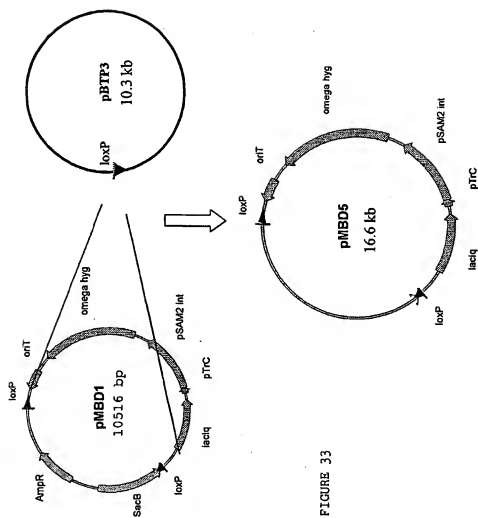


FIGURE 33

35/38

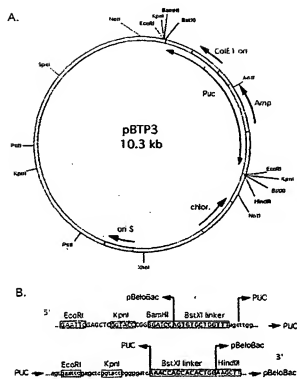


FIGURE 34

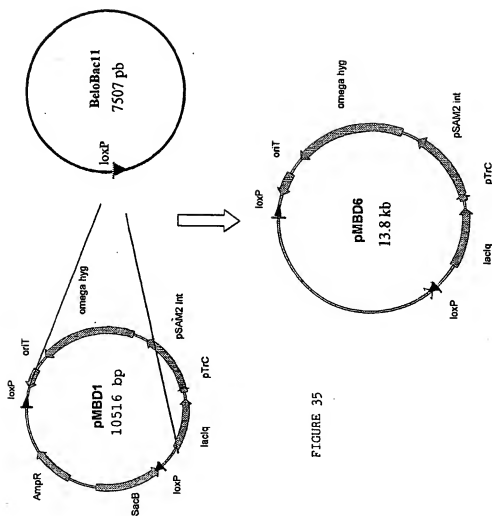


FIGURE 35



37/38

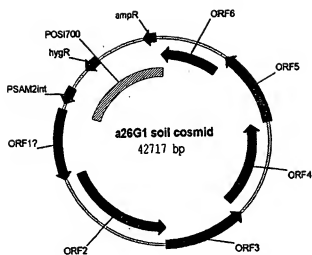


FIGURE 37

38/38

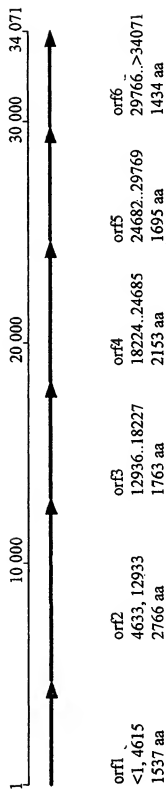


FIGURE 37

## LISTE DE SEQUENCES

<110> Aventis Pharma S.A.

<120> Procédé d'obtention d'acides nucléiques à partir d'un échantillon de sol, acides nucléiques ainsi obtenus et leur application à la synthèse de composés

<130> Banque d'ADN du sol - RPR S.A.

<140>

<141>

<150> FR9915032

<151> 1999-11-29

<160> 126

<170> PatentIn Ver. 2.1

<210> 1

<211> 15

<212> ADN

<213> Séquence artificielle

<220>

<223> Description de la séquence artificielle:sonde  
FGPS431

<220>

<221> variation

<222> (14)

<223> Base A remplacée par G

<400> 1  
acgggcggtg tgtac

<210> 2

<211> 22

<212> ADN

<213> Séquence artificielle

<220>

<223> Description de la séquence artificielle:amorce  
FGPS122

<400> 2  
ggagagtttg atcatggctc ag

<210> 3  
<211> 20  
<212> ADN  
<213> Séquence artificielle

<220>  
<223> Description de la séquence artificielle:amorce  
FGPS350

<400> 3  
cctggagtta agccccaagc

20

<210> 4  
<211> 24  
<212> ADN  
<213> Séquence artificielle

<220>  
<223> Description de la séquence artificielle:sonde  
FGPS643

<220>  
<221> variation  
<222> (20)  
<223> T remplacée par C

<400> 4  
gtgagtnnna acctgcccct gact

24

<210> 5  
<211> 21  
<212> ADN  
<213> Séquence artificielle

<220>  
<223> Description de la séquence artificielle:sonde  
FGPS643-2

<400> 5  
gtgagtaacc tgccccgcac t

21

<210> 6  
<211> 23  
<212> ADN  
<213> Séquence artificielle

<220>  
<223> Description de la séquence artificielle:amorce  
R499

<400> 6  
ttaattcact tgcaactgat ggg

23

<210> 7  
<211> 23  
<212> ADN  
<213> Séquence artificielle

<220>  
<223> Description de la séquence artificielle:amorce  
R500

<400> 7  
aacgatatgct cctacatttg gag

23

<210> 8  
<211> 25  
<212> ADN  
<213> Séquence artificielle

<220>  
<223> Description de la séquence artificielle:sonde C501

<400> 8  
ttgctgatac ggtatagaac ctggc

25

<210> 9  
<211> 20  
<212> ADN  
<213> Séquence artificielle

<220>  
<223> Description de la séquence artificielle:amorce  
FGPS516

<400> 9  
tccagatcct tgaccgcgag

20

<210> 10  
<211> 20  
<212> ADN  
<213> Séquence artificielle

<220>  
<223> Description de la séquence artificielle:amorce  
FGPS517

<400> 10  
cacgacattg cactccaccg 20

<210> 11  
<211> 16  
<212> ADN  
<213> Séquence artificielle

<220>  
<223> Description de la séquence artificielle:sonde  
FGPS518

<400> 11  
ccgtgagccg gatcag 16

<210> 12  
<211> 20  
<212> ADN  
<213> Séquence artificielle

<220>  
<223> Description de la séquence artificielle:FGPS612

<220>  
<221> variation  
<222> (2)  
<223> Base C remplacée par T

<220>  
<221> variation  
<222> (7)  
<223> Base T remplacée par C

<220>  
<221> variation  
<222> (7)  
<223> Base T remplacée par A

<400> 12  
ccaacttcgt gccagagcc 20

<210> 13

<211> 21  
<212> ADN  
<213> Séquence artificielle

<220>  
<223> Description de la séquence artificielle:FGPS669

<220>  
<221> variation  
<222> (7)  
<223> Base A remplacée par G

<220>  
<221> variation  
<222> (13)  
<223> Base A remplacée par C

<400> 13  
gacgtcatcc ccaccttcct c 21

<210> 14  
<211> 18  
<212> ADN  
<213> Séquence artificielle

<220>  
<223> Description de la séquence artificielle:FGPS618

<220>  
<221> variation  
<222> (5)  
<223> Base T remplacée par C

<400> 14  
atggttgctg tcagctcg 18

<210> 15  
<211> 21  
<212> ADN  
<213> Séquence artificielle

<220>  
<223> Description de la séquence artificielle:FGPS614

<400> 15  
gtgtagaagt gaaattcgat t 21

<210> 16  
<211> 18  
<212> ADN  
<213> Séquence artificielle

<220>  
<223> Description de la séquence artificielle:FGPS615

<400> 16  
cggtggatga tgtggatt 18

<210> 17  
<211> 18  
<212> ADN  
<213> Séquence artificielle

<220>  
<223> Description de la séquence artificielle:FGPS616

<400> 17  
aggttaaaac tcaaata 18

<210> 18  
<211> 18  
<212> ADN  
<213> Séquence artificielle

<220>  
<223> Description de la séquence artificielle:FGPS621

<400> 18  
atacgtaggt ggcaagcg 18

<210> 19  
<211> 19  
<212> ADN  
<213> Séquence artificielle

<220>  
<223> Description de la séquence artificielle:FGPS617

<400> 19  
gccgggggtca actcggagg 19

<210> 20  
<211> 18



<212> ADN  
<213> Séquence artificielle  
  
<220>  
<223> Description de la séquence artificielle:FGPS680

<220>  
<221> variation  
<222> (11)  
<223> Base A remplacée par C

<220>  
<221> variation  
<222> (11)  
<223> Base A remplacée par T

<220>  
<221> variation  
<222> (13)  
<223> Base T remplacée par A

<400> 20  
tgagtcccca actccccg 18

<210> 21  
<211> 20  
<212> ADN  
<213> Séquence artificielle

<220>  
<223> Description de la séquence artificielle:FGPS619

<400> 21  
gcttggggct taactccagg 20

<210> 22  
<211> 21  
<212> ADN  
<213> Séquence artificielle

<220>  
<223> Description de la séquence artificielle:amorce 63f

<400> 22  
caggcctaac acatgcaagt c 21

<210> 23

<211> 18  
<212> ADN  
<213> Séquence artificielle

<220>  
<223> Description de la séquence artificielle:amorce  
1387r

<400> 23  
gggcggngtg tacaaggc

18

<210> 24  
<211> 30  
<212> ADN  
<213> Séquence artificielle

<220>  
<223> Description de la séquence artificielle:oligo-1

<400> 24  
gcttatttaa atattaagcg gccgccggg

30

<210> 25  
<211> 28  
<212> ADN  
<213> Séquence artificielle

<220>  
<223> Description de la séquence artificielle:oligo-2

<400> 25  
cccggcggcg cgcattaata tttaaata

28

<210> 26  
<211> 23  
<212> ADN  
<213> Séquence artificielle

<220>  
<223> Description de la séquence artificielle:amorce al

<400> 26  
ccncagnagc gcntnttnc t nga

23

<210> 27  
<211> 22

&lt;212&gt; ADN

&lt;213&gt; Séquence artificielle

&lt;220&gt;

&lt;223&gt; Description de la séquence artificielle:amorce a2

&lt;400&gt; 27

gtncncgtnc cgtgngtntc na

22

&lt;210&gt; 28

&lt;211&gt; 23

&lt;212&gt; ADN

&lt;213&gt; Séquence artificielle

&lt;220&gt;

&lt;223&gt; Description de la séquence artificielle:amorce b1

&lt;400&gt; 28

ccncagnagc gcntnctnct nga

23

&lt;210&gt; 29

&lt;211&gt; 22

&lt;212&gt; ADN

&lt;213&gt; Séquence artificielle

&lt;220&gt;

&lt;223&gt; Description de la séquence artificielle:amorce b2

&lt;400&gt; 29

gtncncgtnc cgtgnccttc na

22

&lt;210&gt; 30

&lt;211&gt; 672

&lt;212&gt; ADN

&lt;213&gt; Streptomyces ambofaciens

&lt;400&gt; 30

cccagcagc acgtgttctt cgagacgggt tgggagacct tcgaatccgc cggagtggac 60  
 ccgcgcgcgg tacgcggctg ttccgtcggg atgttcgtcg gcaccaacgg acaggactac 120  
 ccggtggtgt tggccggatc cgcgcagcag ggcctggacg ccacgcggcg caccggtaac 180  
 gcggcggcgg tgctgtccgg ccgggtctcg tacgccttcg gcctggaagg gccgcggctc 240  
 accgtcgaca cggcgtgttc gtctgcgtg gtggcccttc acctggccgc gcaggcgtg 300  
 cggcgcggcg agtcgatctt ggcactcgcc ggcggtgtgt cggagatgac caccgaggcg 360  
 gcgttcaccg agttcgcccg gcaggggcggc ctggccgacg acggccgctg caaggccttc 420  
 ctggccgacg ccgacggcac gggctggggc gagggcgtcg gcgtcctgct ggtggagcgg 480  
 ttggcggacg cccgcgcgca cgggcacccg gccctcgccg tggtaggggg cagcgcggtc 540  
 aaccaggacg gcgcctccaa cggcttgacg gcacccaacg gcccgctcca gcagcgagtc 600

atccggcagg cactggcgga cgcgcggctg tcgccgtcgg aggtcgacgc ggtcgagacc 660  
cacggcaccc gc 672

<210> 31  
<211> 665  
<212> ADN  
<213> *Streptomyces ambofaciens*

<400> 31  
ccccagcagc gcgtgttcct ggaagcgtcc tgggaggcgg tcgagcgggc aggcacgcac 60  
atgcgcaccc tgcgcgggtg acgcacccgc gtcttcgcgc gcgtgatgta ccacgactac 120  
ccgtcgggtg tcgaccccca agcgcctcgac ggctacctgg gcacggccaa cgcggcgacg 180  
gttctctccg gccgcacatgc ctacaccttc gggcttcagg gaccggcggt caccgtggag 240  
acggcctgct cctcgctccct ggtggcgctg cacctgcgc cccaggcgct gccgcgcggc 300  
gagtgcgaaac tcgcctcgtt cgggtggggtc acggtcatgt ccggcccgat gatgttcgcg 360  
ggcttcggcc tggaagacgg ctctgcgcgc gacggcgctt gcaaggcggt cgcgcgcgcg 420  
gccgacggca cggcctgggg cgagggtgtc ggtgtgctgc tgggtggagcg gctgtcgagc 480  
gcccgcgccg acggggcacc ggtgctggcc gtggtgcgcg gtacgcgggt caaccaggac 540  
ggtgcctccg gcggcctcac cgcgcccaac ggacctgccc agcagcgctt catccgtcag 600  
gcctggcgga gcgcggcact cgtaccggcc gaggtcgacg cggtcgagac ccacggcacc 660  
gggac 665

<210> 32  
<211> 671  
<212> ADN  
<213> *Saccharopolyspora erythraea*

<400> 32  
ccgcaggagc gcgtgttcct ggaactcgct tgggaagcac ttgataacgc gggcatcgca 60  
ccgcacagcc tcagggacag ccggacgggc gtgttcttcg gagctatgtg gcacggctac 120  
gcgcagttcg cagccggagc cgtcgaccgc atcacccagg acaccgcgac cgggacagac 180  
ctgagcatca tcccggccag gatcgcttac ttcttgggct tgcgcggccc ggacatgacc 240  
ctgaacacgc cgtgctcatc ggctttgggt gccatgcacc aggcacgcca aagcatcctg 300  
ctggcgcaat cctcggtcgc cttggtcggc gggtacagct tgttggtcgc gctggagcgc 360  
atggtcgcca tgtcgcgggt cggagcgatg gccccggagc gccgggtgcaa ggcattcgac 420  
tctcgcgcga acggctacgt gcgcgcgcaa ggcggcggtg tcgttgtgtc caaacgcgctg 480  
tcgcgcgctc tggcgcgatg caaccgggtc tactgcgtcc tgcgcggcag cgcggtcaac 540  
aacgcaggct tcagcaatgg ccttacccgc ccgagccggc cggcgcagga gcaggtactg 600  
cgcgacgct acgccaacgc cggggtcgat ccggcacagg tcgactacgt cgagaccacc 660  
gggacggca c 671

<210> 33  
<211> 686  
<212> ADN  
<213> *Organime Inconnu*

<220>

<223> Origine de la séquence :organisme du sol

<400> 33

```

ccgcaggagc gcgtgttctt cgagtcgtgc tgggaggcgc tggagcatgc tggatacgc 60
actgcacgct accccggccg catcgggctg tgggcccggc cgggcttcaa cagctacctc 120
ctgaccaatc tcataaaca cgcgcctttt tttagagagc tgggcatgta ccagatcttt 180
ctgagcaacg acaaggactt catcgccacc cgcacggctt acaagttaa cctgcgcggt 240
ccggcgatgg ccgtcggcac cgcctgttcc acatcgctgg tggcggttca cgaagcttgc 300
caggcgctgc ggtcgggcga gtgtgacatg gcactggcgc gtgctgcgtc tgtcagcacg 360
ccctccggg agggctacct ctaccaggaa ggcatgatta tgagccgtga cggcgtctgc 420
cgcccgctt acgcccacgc ccatggccgc gtgctgggca atggcggtgc ggtcgtggtg 480
ctcaagcgcc tggacgaagc gctccgggac ggtgacacgg tctacgcggt gattcgtggc 540
acggcggtca acaacgacgg ctctgtcaag atcgggttca cggcgccacg cggcgagggg 600
cagagccggg tcgtgcggga cgccctgcgg gcggccgcgg tcccgcggga gagcgtgacc 660
tacgtcgaca cgcacggcac cggcac 686

```

<210> 34

<211> 689

<212> ADN

<213> Organime Inconnu

<220>

<223> Origine de la séquence :organisme du sol

<400> 34

```

ccccagcagc gcctgttctt cgagtcgcgc tgggaagcga tggagaacgc gggatatgcg 60
gcgcgaagct ataagggttc gatcggcgtt ttccggggat ggcgcgtcaa tacctacctg 120
ctgaacaacc tcgccaccgc ggagcgcgtt gattttctac gccctccgc gtaccagctg 180
ctgacggcca acgacaagga ttctctggcc acgcgtgtct cttacaagct gaacctccgc 240
gggcccagcc tgacggttca gacggcgtgc tccacctcgc tgggtgtcgtt ggtgatggca 300
tgcgagagct tgcagcgccg cgcctcggac attgccttgg cggggggagt tgccatcaat 360
gttcgcgagt ccgtggggta cctgcaccag cggggcatga tcctgtgcgc cgaacggcgc 420
tgccgcgctt tcgatgagtc cgtcaaggc acggtgccgg gcaacggcgc ggggtgtggtc 480
gtctcgaagc gcttgagccg cgtctgtgcc gatggcgaca ccatctacgc cgtcattcgc 540
ggagcgggta ttaataatga tggcgccgag cgcattgggt ttacogctcc aggtgtggac 600
ggtcagacgc gattgattgc gcgactcaa gagatggcgg gcgtgaagcc ggagtcacatc 660
ggctacatgg acaccacgag caccggcac 689

```

<210> 35

<211> 671

<212> ADN

<213> Organime Inconnu

<220>

<223> Origine de la séquence :organisme du sol

<400> 35

```

ccgcagcagc gcctcttctt cgaggtggca tgggaagctt tggagcgtgc gggtcggccg 60

```

```

cccacagctc tcgcgggcgag cgacacggga gtgttcacgc ggatcagcac cgacgactac 120
agccggctga aacctacoga tccggcgctc attgacgcct ataccggtac cggaacgcgc 180
ttcagcactg ccgccggagc gatctcctat ctgctggggg tgcaggggacc gaacttcccc 240
gtcgacacgg cgtgctcttc ctcaactcgt gcggttcate tggcggtgcc cagcttgacg 300
tcgcgagagt gcagcatggc gctggcgggc ggcgtagaac tgattctggc gccggaaagc 360
acgatctact tctgcgcctt cggggccatg gcggccgatg gccgttgcaa aagtttcgct 420
cgctccggcg acggttacgg ccgcgccgag ggaatgcggaa tgcgtgtgct gaagcggctg 480
tccgatgcga cgcgtgacgg cgatcgtatt ctggcgctga ttcggggatc ggccgtcaac 540
cacggcgggc gcagcaacgg cctcacggcg ccgaacggtc cggcgaggga agcgtgatt 600
cggcgggcgc tcaagaacgc cggcatggcc ccgcccgatg tcgattacgt ggacacccac 660
ggcacccgga c 671

```

&lt;210&gt; 36

&lt;211&gt; 758

&lt;212&gt; ADN

&lt;213&gt; Organisme Inconnu

&lt;220&gt;

&lt;223&gt; Origine de la séquence :organisme du sol

&lt;400&gt; 36

```

ccgcaggagc gcgtcttctt cgaacgcatt gacggtttgc atcggaattt cttcggcatc 60
tccccccgcg aagctctgaa catggatccg cagcagcgcc tgctgctgga agtggtctgg 120
gaagcggcag aggaacgcgc catctctccc ggccctctgg ccggcagcgc gaccggcgct 180
tttgcgggct cctgcgccca ggacttcgga ctgtttcagt acgcgcgacc tgcccgcac 240
ggagcttggt cgggttccgc cgtggcgcat agcatgttg ccaatcgcat ctctatctg 300
ctcgacctgc gcggtccgag catggcggtc gatacggcct gctcctccg gctcgtgcc 360
gtccatctgg cttgccaaag cctgcgcggc cgcgaatgca atgcggcatt cgcggcgga 420
gtgaacttga tcttgactcc cgagggcatg atcgctttg cgaaggctcg catgttgagg 480
cccgcaggac gctgcaagac gttcgacgcc gcagccgacg gttatgtgc cggcgagggc 540
tgcgggcatc tgctgctgaa cgggctctcc gatgcgctgg ccgatggcga tgccatctgt 600
cgagtcaccc cgggctcggc aatcaatcag gacggacgga gcaatggcat caccggcgcc 660
aatctgcagg gcagaaaggc ggtcctgcaa gaggcggtgg ccaacgcgca catcgatcca 720
tcccacgtat cgttgatcga caccgacggc accgcgac 758

```

&lt;210&gt; 37

&lt;211&gt; 704

&lt;212&gt; ADN

&lt;213&gt; Organisme Inconnu

&lt;220&gt;

&lt;223&gt; Origine de la séquence :organisme du sol

&lt;400&gt; 37

```

ccgcagcagc gcgtgttctt cagatgcgcc tgggaggcgg tggaaagcgc gggctacagt 60
cccgaaaaat atcccggcct gatcggagtt ttccgcgggg ccagcatcaa cagctatttc 120
ctttataacc tcgcgcacaa ccgggaattc gtgcgccgca tggcggggga gtaccaagt 180
ggcgagtacc agacgatcct cggaaacgac aaggactacc tcccactcgc cgtctccctac 240

```

```

aaattgaacc tgcgcggccc cagcctggcc gtgcagtcgc cctgctcgac cggcctcgtc 300
gccgtttgtc aggccattca aaatctgcag acttatcagt gcgatatggc cctcgcgggc 360
ggcatctcga ttctgtttcc gcaaaagcgc gactaccgct tcaccgacga aggaatggtc 420
tctcgcgacg gtcactgcgc ccggttcgac gccagcgcgc aaggcacggt cttcgcgcaac 480
ggggccggcg tcgtcctgat gaaaagattg gccgacgcag tgaccgatcg ggacacgato 540
ctcgcctgtg ttagggggcgc tgccgtgaac aacgacggcg gcgtcaaaaat gggttacacg 600
gcgccacgtg ccgaaggtca ggcggaggcc atcaccctgg ccctcgcgct cgctggcgctc 660
agcccgagga ccatcacttg catggacacc cagggcacgc gcac 704

```

&lt;210&gt; 38

&lt;211&gt; 680

&lt;212&gt; ADN

&lt;213&gt; Organisme Inconnu

&lt;220&gt;

&lt;223&gt; Origine de la séquence :organisme du sol

&lt;400&gt; 38

```

ccccagcagc gcgtgttctt cgaatgcgcc tgggcggcgc tggagcgccg ccggatatca 60
gggcgacacc ttccacgggtg tccatcgccg gtctatgcct caagcggctt taacacctat 120
cttctgaacc tgcattgcaa tgccgcgggtg ccgcaatcga tcagcccggt tgaactgttc 180
gtcgccaaag acaaggattt tctggcgacg cgcacggctt acaagctcaa tctgcgcggc 240
ccggccatga cagtgcagac ggctctgctc tcactgttgg ttgccgttca tgcgcgcgcg 300
caaagcctcc tagcggggga atgcgatatt gcgctcgcgg gcggcatcac ggtttccctg 360
tcgcatggat atgtggcgcg cgaaggtgga atattgtctc ctgacgggca ttgccggggc 420
ttcgatcgcg atgcggcgcg aaccgttcca ggcagcggcg tcggcggtgt cgtgctcaag 480
cgtctcgaag atgcgcttgc agacggcgat acgatcgacg cgtcatcat cggttcgccc 540
atcaacaatg atggcgcgct gaaggcgagc tttaaccgac cgcagggtga cagccaggcc 600
ttggtcatca gcgagggcca tcgactgccc ggaatatcgg ccgattccat cggttatatg 660
gacacccacg gcaccgggac 680

```

&lt;210&gt; 39

&lt;211&gt; 671

&lt;212&gt; ADN

&lt;213&gt; Organisme Inconnu

&lt;220&gt;

&lt;223&gt; Origine de la séquence :organisme du sol

&lt;400&gt; 39

```

ccgcagcagc gcctcttctt cgaagtcacc tgggaagcgc tgggaagatgc cggcatcccc 60
ccgtccacga ttgcggcgac gaatgtcggc gttttctatgg gcgcgtcgca ggctgactac 120
ggccacaagt tcttcagcga ccaagccgct cgggattccc atttcgccac cggcacctcg 180
ctggcggttg tcgccaatcg catttctcac atctacgacc tgcgcggccc aagcctcaact 240
gtagacacgc cgtgctcgtc gtcgctcgtc gcgctgcac agcggttggga agcgtctcgc 300
tcggggcgga tcgaaacagc cattgtcggc ggcattaacg ttatcgccag ccggcgctcc 360
ttcatcgctt tctcgcagcg ctcgatgctg tcgccgacgg ggttggtcca ggtcttctcc 420
gccaaaggcg atggctttgt ccgcggcgag ggcggcacgg ttttctctct gcgcaaggcg 480

```

```

gcgcatcgcg atggcagccg caaccgggtg cgcgggctca ttctcgccac cgacgtcaat 540
tccgacgggc gtaccaacgg catctcgctg ccatcgcccg aagcgaggga agtctctctg 600
caacgcgtct attcacgcgc atcgatcgat ccgaaccgac tggctttctg cgacacccac 660
gggaccggca c 671

```

&lt;210&gt; 40

&lt;211&gt; 764

&lt;212&gt; ADN

&lt;213&gt; Organisme Inconnu

&lt;220&gt;

&lt;223&gt; Origine de la séquence :organisme du sol

&lt;400&gt; 40

```

ccgcagcagc gcgtgttcct cgacggcatc gaccggttcg atccgcgtca ctctcgcatc 60
acgcccgcgc aggcgatcag catggaccgc cagcagcggc tcctgtctga gggtcagctg 120
gaagcgctgg agcgcgcgcg cgtggcgccc gatcgctga ccggatccga caccggcgctc 180
ttcatcgcca tcagcaccac cgactacggc cagatcctg tcgcgcgcctc ggaccagctc 240
gatccgggga tgaacttcgg caccggcaac ctgttgaacg cgcgcgcgcg acgcctctcg 300
tacgtctctg gcctgcaggg tccgagcatg cgggtcgaca ccgcattgtcc gtcgtcgctg 360
gtggcgatcc atctcgctg tcagagcctg cgcaaccgcg agtgcgcgat ggcgtctcgcc 420
ggcggcgcca acctgggtgt cgtcccgga gtgacggta actgctgcgc cgccaagatg 480
ctcgcgcctg acggcgctg caagacgttc gacgcgcgcg cggacggcta cgtccgcgcg 540
gaagggggcc cggtgatcgt gctgaagcgc ctctccgacg cgctggcgga cggcgatccg 600
atcgtcgcg tgatccgcgc atccgcggtc aatcaggacg gccgcagcgc cggcttcacc 660
gcgcgcaacg aactggcgca gcaggcggtg atccggaccg cgctcgcgcc agcggcgctc 720
gcgcgctccg acatcggtca cgtggacacg caccggaccg ggac 764

```

&lt;210&gt; 41

&lt;211&gt; 763

&lt;212&gt; ADN

&lt;213&gt; Organisme Inconnu

&lt;220&gt;

&lt;223&gt; Origine de la séquence :organisme du sol

&lt;400&gt; 41

```

ccgcagcagc gcgtgttcct cgacggcatc gaccgcttcg atccgcagtt ttccgggcatc 60
gcgcgcgcgc aagcggccgc catcgatccg cagcagcggc tgctgtctga ggcagcagctg 120
gaagcgctgg aagacgcgcg gacgtcgccc aggaaccccc gcgcggcgctg 180
ttcgtcgcca tcaacagcat cgactacgcg acgctgcagc tcgagaactg cgtatctggcc 240
agcatcgagc cctattcgct ctccggcagc gcgcacagca tcgcggcgcc ggcgtctgcc 300
tacgtgtctg gcctgcaggg gccggcgatg gcggctcgaca ccgcctgtct gtcgtcgctg 360
gtcgcgatcc acctggcgtg ccagagcctg cgcaaccgac actgcgcgct gcgcgtggcc 420
ggcggcgctg acgtcacgct gacgccgata aacatggtcg tgttctcgaa gctcgcatg 480
ctggcgcgcc acggcaagtg caagacgttc gacggccgcg gcgacggatt cgtcgaaggc 540
gagggtctcg cggtcatcgt cctcaagcgc ttgtcgacg cgcttgccga caaggatcgg 600
atcctcgcg tggtgcgcgg ttccggcggtc aaccaggacg gcgcgagcag cgtctctacc 660

```



gcgcgaacg gtccggcgca ggaagcggtc atccgcgcgg cgttgaagcg ggccggcggtg 720  
cagccggcgg aggtcgcgta cgtggacacc caccggaccc gca 763

<210> 42  
<211> 668  
<212> ADN  
<213> Organisme Inconnu

<220>  
<223> Origine de la séquence :organisme du sol

<400> 42  
ccgcaggagc gcgtgctgct ggaatcctcg tggcatgcgc tggaaagacg cggctatgcc 60  
ggcgaaagca tcgccggcgc gcgctgcggc gtgtacatgg gcttcaacgg cggcgactac 120  
ggcgacctgc tgtacggcca gccctcgctg ccgcgcacg cgaatgtggg caacgccgcc 180  
tcggtgctgt cggcgcgcat cgcctattac ctggacctgc aaggcccgcc gatcacctc 240  
gacaccgctt gttcgagctc gttggctcgg gtgcattctgg cctgccaggg gctgtggacc 300  
ggcgagaccg atctggccct ggccggcgcg gtgtggatcc agtgcaacgc cggattcctg 360  
atctcctcca gccgcgcggc catgctctcg ccgaccggcc agtgccgcgc gttcggcgcc 420  
ggcgccgacg gcttcgtgcc gtccgaaggc gtccggcggtg tcgtgctcaa gcgctgcag 480  
gacgcgctcg acgcggcgca ccacatntac ggctgatcc gcggcagcgc gatcaaccag 540  
gacggcgcca gcaacggcat caccgcgcgg agcgcgcggc ccaggagcg cttgcagcgc 600  
cacgtctacg acagcttcgg catcgacgcc tcgcgcctgc agatgatcga ggcccaacggc 660  
accggcac 668

<210> 43  
<211> 671  
<212> ADN  
<213> Organisme Inconnu

<220>  
<223> Origine de la séquence :organisme du sol

<400> 43  
ccgcaggagc gcgtgctgct ggaggtgact tgggaggcac tcgaagacgc cggccaagac 60  
gtggaccgtc tggccggggc gcccgctcgc gtcttcgtcg ggatctcgtc gaacgattac 120  
ggccagcttc agaacggcga cccggccgac gtggacgcct acgtcggcac cggtaacgcg 180  
ctgagcatcg ccgccaaacg actcagctac acgtttgact ttccggggcc gactctggcg 240  
gtggacacgg cgtgctcgtc ttcactcgtc gcgatccatc tcgcctgccca gacggttcgc 300  
cgcggtgaag cggaaactgc cgtcgcggcc ggctcaact tgattctgac cccggcgctg 360  
acggtgaatt tcacccgcgc cggcatgatg gcgcctgacg gccggtgcaa cagcttcgac 420  
gcggcgccga acggctacgt gcgcggcgaa ggccgcggcg tcgtcgtgct caagccgctg 480  
gcccaggcta tcgcccagcg cgacccgcatc tacgcgatcg tcggtggcag cgcgctcaac 540  
caggacggcc gttccaacgg cctcaccgcc ccgaaccgac agggcccaaga ggtcgtgctg 600  
cgggcgcgct atcgtgacgc gggcatcagc ccggccgatg tcgacgcgct caggcccaac 660  
ggaccggca c 671

<210> 44  
 <211> 707  
 <212> ADN  
 <213> Organisme Inconnu

<220>  
 <223> Origine de la séquence :organisme du sol

<400> 44  
 ccccagcagc gcgtgttcct cgaggacgcg actgaggtcg acgtggatgc gctttcagac 60  
 ggcgaaagacg tcgtgatcgc cggtcatcatg cagcacatcg aggaggccgg catccactcg 120  
 ggcgattcat cgtgcgtgct tccgccggtc gacatcccgc cgaaggcgct gcagacgatc 180  
 cgcgatcaca cgttcaagct cgcgcgcgcg ttgaaggta cgggctgat gaacgtgcag 240  
 tacgcgattc agcgcgacaa ggtctacgtg attgaggtaa accctagggc ttctcgaact 300  
 gtcccgtatg tctcgaaggc gacaggcgtg ccgctggcga aggtcgcgtc acgcttgatg 360  
 accggagcga aactgcacga gctgttccg gcaggggctg agcgcggctg gatcaccacc 420  
 gcggggcgaga atttctacgt gaagtcgccg gtcttcccgt ggggtaagt cccgggctgt 480  
 gacactgtgc tcggggcgga gatgaaatcg accggcggaag tcatgggctg cggcgacaac 540  
 ttcggcgagg ccttcgcaaa ggcacagatc gccgccggca catacctgcc gaccgaaggt 600  
 accgtcttca tctcgggtcaa cgaccgtgac aaaggcaacg tcattcagct ggcgcagcgt 660  
 ttctccgaac tcggtttcgg cattgtcgac acgcacggca ccgggac 707

<210> 45  
 <211> 225  
 <212> PRT  
 <213> Streptomyces ambofaciens

<400> 45  
 Pro Gln Gln His Val Phe Leu Glu Thr Val Trp Glu Thr Phe Glu Ser  
 1 5 10 15  
 Ala Gly Val Asp Pro Arg Ala Val Arg Gly Arg Ser Val Gly Met Phe  
 20 25 30  
 Val Gly Thr Asn Gly Gln Asp Tyr Pro Val Val Leu Ala Gly Ser Ala  
 35 40 45  
 Asp Glu Gly Leu Asp Ala His Ala Ala Thr Gly Asn Ala Ala Ala Val  
 50 55 60  
 Leu Ser Gly Arg Val Ser Tyr Ala Phe Gly Leu Glu Gly Pro Ala Val  
 65 70 75 80  
 Thr Val Asp Thr Ala Cys Ser Ser Ser Leu Val Ala Leu His Leu Ala  
 85 90 95  
 Ala Gln Ala Leu Arg Arg Gly Glu Cys Asp Leu Ala Leu Ala Gly Gly  
 100 105 110

Val Ser Glu Met Ser Thr Glu Ala Ala Phe Thr Glu Phe Ala Arg Gln  
 115 120 125

Gly Gly Leu Ala Asp Asp Gly Arg Cys Lys Ala Phe Ser Ala Asp Ala  
 130 135 140

Asp Gly Thr Gly Trp Gly Glu Gly Val Gly Val Leu Leu Val Glu Arg  
 145 150 155 160

Leu Ala Asp Ala Arg Arg Asn Gly His Arg Ala Leu Ala Leu Val Arg  
 165 170 175

Gly Ser Ala Val Asn Gln Asp Gly Ala Ser Asn Gly Leu Thr Ala Pro  
 180 185 190

Asn Gly Pro Ser Gln Gln Arg Val Ile Arg Gln Ala Leu Ala Asp Ala  
 195 200 205

Arg Leu Ser Pro Ser Glu Val Asp Ala Val Glu Thr His Gly Thr Gly  
 210 215 220

Thr  
 225

<210> 46

<211> 207

<212> PRT

<213> *Streptomyces ambofaciens*

<400> 46

Ala Ser Trp Glu Ala Val Glu Arg Ala Gly Ile Asp Met Arg Thr Leu  
 1 5 10 15

Arg Gly Gly Arg Thr Gly Val Phe Ala Gly Val Met Tyr His Asp Tyr  
 20 25 30

Pro Ser Val Val Asp Pro Glu Ala Leu Asp Gly Tyr Leu Gly Thr Ala  
 35 40 45

Asn Ala Gly Ser Val Leu Ser Gly Arg Ile Ala Tyr Thr Phe Gly Leu  
 50 55 60

Gln Gly Pro Ala Val Thr Val Asp Thr Ala Cys Ser Ser Ser Leu Val  
 65 70 75 80

Ala Leu His Leu Ala Ala Gln Ala Leu Pro Ala Gly Glu Cys Glu Leu  
 85 90 95

Ala Leu Val Gly Gly Val Thr Val Met Ser Gly Pro Met Met Phe Ala  
100 105 110

Gly Phe Gly Leu Glu Asp Gly Ser Ala Ala Asp Gly Arg Cys Lys Ala  
115 120 125

Phe Ala Ala Ala Ala Asp Gly Thr Gly Trp Gly Glu Gly Val Gly Val  
130 135 140

Leu Leu Val Glu Arg Leu Ser Asp Ala Arg Arg His Gly His Arg Val  
145 150 155 160

Leu Ala Val Val Arg Gly Ser Ala Val Asn Gln Asp Gly Ala Ser Gly  
165 170 175

Gly Leu Thr Ala Pro Asn Gly Pro Ala Gln Gln Arg Val Ile Arg Gln  
180 185 190

Ala Leu Ala Ser Ala Ala Leu Val Pro Ala Glu Val Asp Ala Val  
195 200 205

<210> 47

<211> 223

<212> PRT

<213> *Saccharopolyspora erythraea*

<400> 47

Pro Gln Glu Arg Val Phe Leu Glu Leu Ala Trp Glu Ala Leu Asp Asn  
1 5 10 15

Ala Gly Ile Ala Pro His Ser Leu Arg Asp Ser Arg Thr Gly Val Phe  
20 25 30

Phe Gly Ala Met Trp His Gly Tyr Ala Gln Phe Ala Ala Gly Ala Val  
35 40 45

Asp Arg Ile Thr Gln His Thr Ala Thr Gly His Asp Leu Ser Ile Ile  
50 55 60

Pro Ala Arg Ile Ala Tyr Phe Leu Gly Leu Arg Gly Pro Asp Met Thr  
65 70 75 80

Leu Asn Thr Ala Cys Ser Ser Ala Leu Val Ala Met His Gln Ala Arg  
85 90 95

Gln Ser Ile Leu Leu Gly Glu Ser Ser Val Ala Leu Val Gly Gly Ile  
100 105 110

Ser Leu Leu Val Ala Leu Asp Ser Met Val Ala Met Ser Arg Phe Gly  
 115 120 125  
 Ala Met Ala Pro Asp Gly Arg Cys Lys Ala Phe Asp Ser Arg Ala Asn  
 130 135 140  
 Gly Tyr Val Arg Gly Glu Gly Gly Gly Val Val Leu Lys Pro Leu  
 145 150 155 160  
 Ser Arg Ala Leu Ala Asp Gly Asn Pro Val Tyr Cys Val Leu Arg Gly  
 165 170 175  
 Ser Ala Val Asn Asn Asp Gly Phe Ser Asn Gly Leu Thr Ala Pro Ser  
 180 185 190  
 Pro Ala Ala Gln Glu Gln Val Leu Arg Asp Ala Tyr Ala Asn Ala Gly  
 195 200 205  
 Val Asp Pro Ala Gln Val Asp Tyr Val Glu Thr His Gly Thr Gly  
 210 215 220

&lt;210&gt; 48

&lt;211&gt; 211

&lt;212&gt; PRT

&lt;213&gt; Organisme Inconnu

&lt;220&gt;

&lt;223&gt; Origine de la séquence :organisme du sol

&lt;400&gt; 48

Ser Cys Trp Glu Ala Leu Glu His Ala Gly Tyr Asp Thr Ala Arg Tyr  
 1 5 10 15  
 Pro Gly Arg Ile Gly Leu Trp Ala Gly Ala Gly Phe Asn Ser Tyr Leu  
 20 25 30  
 Leu Thr Asn Leu Met Asn Asn Arg Ala Phe Leu Glu Ser Val Gly Met  
 35 40 45  
 Tyr Gln Ile Phe Leu Ser Asn Asp Lys Asp Phe Ile Ala Thr Arg Thr  
 50 55 60  
 Ala Tyr Lys Leu Asn Leu Arg Gly Pro Ala Met Ala Val Gly Thr Ala  
 65 70 75 80  
 Cys Ser Thr Ser Leu Val Ala Val His Glu Ala Cys Gln Ala Leu Arg  
 85 90 95

```

Leu Gly Glu Cys Asp Met Ala Leu Ala Gly Ala Ala Ser Val Ser Thr
 100 105 110

Pro Leu Arg Glu Gly Tyr Leu Tyr Gln Glu Gly Met Ile Met Ser Arg
 115 120 125

Asp Gly Val Cys Arg Pro Phe Asp Ala Asp Ala Asp Gly Thr Val Leu
 130 135 140

Gly Asn Gly Val Ala Val Val Val Leu Lys Arg Leu Asp Glu Ala Leu
 145 150 155 160

Arg Asp Gly Asp Thr Val Tyr Ala Val Ile Arg Gly Thr Ala Val Asn
 165 170 175

Asn Asp Gly Ser Val Lys Ile Gly Phe Thr Ala Pro Ser Ala Glu Gly
 180 185 190

Gln Ser Arg Val Val Arg Asp Ala Leu Arg Ala Ala Ala Val Pro Ala
 195 200 205

Glu Ser Val
 210

```

<210> 49

&lt;211&gt; 229

&lt;212&gt; PRT

<213> Organime Inconnu

**<220>**

<223> Origine de la séquence : organisme du sol

&lt;400&gt; 49

Pro Gln Gln Arg Leu Phe Leu Glu Cys Ala Trp Glu Ala Met Glu Asn  
1 5 10 15

Ala Gly Tyr Ala Ala Arg Ser Tyr Lys Gly Ser Ile Gly Val Phe Ala  
20 25 30

Gly Cys Gly Val Asn Thr Tyr Leu Leu Asn Asn Leu Ala Thr Ala Glu  
35 40 45

Pro Phe Asp Phe Ser Arg Pro Ser Ala Tyr Gln Leu Leu Thr Ala Asn  
50 55 60

Asp Lys Asp Phe Leu Ala Thr Arg Val Ser Tyr Lys Leu Asn Leu Arg  
65 70 75 80

Gly Pro Ser Leu Thr Val Gln Thr Ala Cys Ser Thr Ser Leu Val Ser  
                     85                    90                    95  
 Val Val Met Ala Cys Glu Ser Leu Gln Arg Gly Ala Ser Asp Ile Ala  
                     100                    105                    110  
 Leu Ala Gly Gly Val Ala Ile Asn Val Pro Gln Ser Val Gly Tyr Leu  
                     115                    120                    125  
 His Gln Pro Gly Met Ile Leu Ser Pro Asp Gly Arg Cys Arg Ala Phe  
                     130                    135                    140  
 Asp Glu Ser Ala Gln Gly Thr Val Pro Gly Asn Gly Ala Gly Val Val  
                     145                    150                    155                    160  
 Val Leu Lys Arg Leu Ser Arg Ala Leu Ala Asp Gly Asp Thr Ile Tyr  
                     165                    170                    175  
 Ala Val Ile Arg Gly Ala Ala Ile Asn Asn Asp Gly Ala Glu Arg Met  
                     180                    185                    190  
 Gly Phe Thr Ala Pro Gly Val Asp Gly Gln Thr Arg Leu Ile Arg Arg  
                     195                    200                    205  
 Thr Gln Glu Met Ala Gly Val Lys Pro Glu Ser Ile Gly Tyr Met Asp  
                     210                    215                    220  
 Thr His Gly Thr Gly  
 225

<210> 50  
 <211> 223  
 <212> PRT  
 <213> Organisme. Inconnu

<220>  
 <223> Origine de la séquence :organisme du sol

<400> 50  
 Pro Gln Gln Arg Leu Phe Leu Glu Val Ala Trp Glu Ala Leu Glu Arg  
                     1                    5                    10                    15  
 Ala Gly Arg Pro Pro Asp Ser Leu Ala Gly Ser Asp Thr Gly Val Phe  
                     20                    25                    30  
 Ile Gly Ile Ser Thr Asp Asp Tyr Ser Arg Leu Lys Pro Thr Asp Pro  
                     35                    40                    45

Ala Leu Ile Asp Ala Tyr Thr Gly Thr Gly Thr Ala Phe Ser Thr Ala  
 50 55 60

Ala Gly Arg Ile Ser Tyr Leu Leu Gly Leu Gln Gly Pro Asn Phe Pro  
 65 70 75 80

Val Asp Thr Ala Cys Ser Ser Ser Leu Val Ala Val His Leu Ala Cys  
 85 90 95

Arg Ser Leu Gln Ser Arg Glu Cys Ser Met Ala Leu Ala Gly Gly Val  
 100 105 110

Asn Leu Ile Leu Ala Pro Glu Ser Thr Ile Tyr Phe Cys Arg Leu Arg  
 115 120 125

Ala Met Ala Ala Asp Gly Arg Cys Lys Ser Phe Ala Ala Ser Ala Asp  
 130 135 140

Gly Tyr Gly Arg Gly Glu Gly Cys Gly Met Leu Val Leu Lys Arg Leu  
 145 150 155 160

Ser Asp Ala Thr Arg Asp Gly Asp Arg Ile Leu Ala Leu Ile Arg Gly  
 165 170 175

Ser Ala Val Asn His Gly Gly Arg Ser Asn Gly Leu Thr Ala Pro Asn  
 180 185 190

Gly Pro Ala Gln Glu Ala Val Ile Arg Ala Ala Leu Lys Asn Ala Gly  
 195 200 205

Met Ala Pro Ala Asp Val Asp Tyr Val Asp Thr His Gly Thr Gly  
 210 215 220

&lt;210&gt; 51

&lt;211&gt; 252

&lt;212&gt; PRT

&lt;213&gt; Organisme Inconnu

&lt;220&gt;

&lt;223&gt; Origine de la séquence :organisme du sol

&lt;400&gt; 51

Pro Gln Glu Arg Val Phe Leu Glu Arg Ile Asp Gly Phe Asp Ala Glu  
 1 5 10 15

Phe Phe Gly Ile Ser Pro Arg Glu Ala Leu Asn Met Asp Pro Gln Gln  
 20 25 30



Arg Leu Leu Leu Glu Val Cys Trp Glu Ala Ala Glu Asp Ala Gly Ile  
           35                    40                    45  
 Ser Pro Gly Pro Leu Ala Gly Ser Ala Thr Gly Val Phe Ala Gly Ser  
           50                    55                    60  
 Cys Ala Gln Asp Phe Gly Leu Phe Gln Tyr Ala Asp Pro Ala Arg Ile  
           65                    70                    75                    80  
 Gly Ala Trp Ser Gly Ser Gly Val Ala His Ser Met Leu Ala Asn Arg  
                     85                    90                    95  
 Ile Ser Tyr Leu Leu Asp Leu Arg Gly Pro Ser Met Ala Val Asp Thr  
                     100                    105                    110  
 Ala Cys Ser Ser Ala Leu Val Ala Val His Leu Ala Cys Gln Ser Leu  
                     115                    120                    125  
 Arg Arg Arg Glu Cys Asp Ala Ala Phe Ala Gly Gly Val Asn Leu Ile  
           130                    135                    140  
 Leu Thr Pro Glu Gly Met Ile Ala Leu Ser Lys Ala Arg Met Leu Ala  
           145                    150                    155                    160  
 Pro Asp Gly Arg Cys Lys Thr Phe Asp Ala Ala Ala Asp Gly Tyr Val  
                     165                    170                    175  
 Arg Gly Glu Gly Cys Gly Ile Val Leu Leu Lys Arg Leu Ser Asp Ala  
                     180                    185                    190  
 Leu Ala Asp Gly Asp Ala Ile Cys Ala Val Ile Arg Gly Ser Ala Ile  
                     195                    200                    205  
 Asn Gln Asp Gly Arg Ser Asn Gly Ile Thr Ala Pro Asn Leu Gln Ala  
           210                    215                    220  
 Gln Lys Ala Val Leu Gln Glu Ala Val Ala Asn Ala His Ile Asp Pro  
           225                    230                    235                    240  
 Ser His Val Ser Leu Ile Asp Thr His Gly Thr Gly  
                     245                    250

&lt;210&gt; 52

&lt;211&gt; 234

&lt;212&gt; PRT

&lt;213&gt; Organime Inconnu

&lt;220&gt;

<223> Origine de la séquence : organisme du sol

<400> 52

```

Pro Gln Gln Arg Val Phe Leu Glu Cys Ala Trp Glu Ala Val Glu Ser
 1 5 10 15

Ala Gly Tyr Asp Pro Glu Lys Tyr Pro Gly Leu Ile Gly Val Phe Ala
 20 25 30

Gly Ala Ser Ile Asn Ser Tyr Phe Leu Tyr Asn Leu Ala His Asn Arg
 35 40 45

Glu Phe Val Ala Arg Met Ala Gly Glu Tyr Gln Val Gly Glu Tyr Gln
 50 55 60

Thr Ile Leu Gly Asn Asp Lys Asp Tyr Leu Pro Thr Arg Val Ser Tyr
 65 70 75 80

Lys Leu Asn Leu Arg Gly Pro Ser Leu Ala Val Gln Ser Ala Cys Ser
 85 90 95

Thr Gly Leu Val Ala Val Cys Gln Ala Ile Gln Asn Leu Gln Thr Tyr
 100 105 110

Gln Cys Asp Met Ala Leu Ala Gly Gly Ile Ser Ile Ser Phe Pro Gln
 115 120 125

Lys Arg Asp Tyr Arg Phe Thr Asp Glu Gly Met Val Ser Arg Asp Gly
 130 135 140

His Cys Arg Pro Phe Asp Ala Ser Ala Gln Gly Thr Val Phe Gly Asn
 145 150 155 160

Gly Ala Gly Val Val Leu Met Lys Arg Leu Ala Asp Ala Val Thr Asp
 165 170 175

Arg Asp Thr Ile Leu Ala Val Ile Arg Gly Ala Ala Val Asn Asn Asp
 180 185 190

Gly Gly Val Lys Met Gly Tyr Thr Ala Pro Ser Ala Glu Gly Gln Ala
 195 200 205

Glu Ala Ile Thr Leu Ala Leu Ala Leu Ala Gly Val Ser Pro Glu Thr
 210 215 220

Ile Thr Cys Met Asp Thr His Gly Thr Gly
 225 230

```

<210> 53  
 <211> 226  
 <212> PRT  
 <213> Organisme Inconnu

<220>  
 <223> Origine de la séquence :organisme du sol

<400> 53  
 Pro Gln Gln Arg Val Phe Leu Glu Cys Ala Trp Ala Ala Leu Glu Arg  
 1 5 10 15

Arg Arg Ile Ser Gly Arg His Leu Pro Arg Cys Pro Ser Ala Val Tyr  
 20 25 30

Ala Ser Ser Gly Phe Asn Thr Tyr Leu Leu Asn Leu His Ala Asn Ala  
 35 40 45

Ala Val Arg Gln Ser Ile Ser Pro Phe Glu Leu Phe Val Ala Asn Asp  
 50 55 60

Lys Asp Phe Leu Ala Thr Arg Thr Ala Tyr Lys Leu Asn Leu Arg Gly  
 65 70 75 80

Pro Ala Met Thr Val Gln Thr Ala Cys Ser Ser Ser Leu Val Ala Val  
 85 90 95

His Val Ala Ala Gln Ser Leu Leu Ala Gly Glu Cys Asp Ile Ala Leu  
 100 105 110

Ala Gly Gly Ile Thr Val Ser Arg Ser His Gly Tyr Val Ala Arg Glu  
 115 120 125

Gly Gly Ile Leu Ser Pro Asp Gly His Cys Arg Ala Phe Asp Ala Asp  
 130 135 140

Ala Gly Gly Thr Val Pro Gly Ser Gly Val Gly Val Val Val Leu Lys  
 145 150 155 160

Arg Leu Glu Asp Ala Leu Ala Asp Gly Asp Thr Ile Asp Ala Val Ile  
 165 170 175

Ile Gly Ser Ala Ile Asn Asn Asp Gly Ala Leu Lys Ala Ser Phe Thr  
 180 185 190

Ala Pro Gln Val Asp Ser Gln Ala Leu Val Ile Ser Glu Ala His Ala  
 195 200 205

Ala Ala Gly Ile Ser Ala Asp Ser Ile Gly Tyr Met Asp Thr His Gly  
 210 215 220

Thr Gly  
225

<210> 54  
<211> 223  
<212> PRT  
<213> Organisme Inconnu

<220>  
<223> Origine de la séquence : organisme du sol

<400> 54  
Pro Gln Gln Arg Leu Phe Leu Glu Leu Thr Trp Glu Ala Leu Glu Asp  
1 5 10 15  
Ala Gly Ile Pro Pro Ser Thr Ile Ala Gly Thr Asn Val Gly Val Phe  
20 25 30  
Met Gly Ala Ser Gln Ala Asp Tyr Gly His Lys Phe Phe Ser Asp His  
35 40 45  
Ala Val Ala Asp Ser His Phe Ala Thr Gly Thr Ser Leu Ala Val Val  
50 55 60  
Ala Asn Arg Ile Ser Tyr Ile Tyr Asp Leu Arg Gly Pro Ser Leu Thr  
65 70 75 80  
Val Asp Thr Ala Cys Ser Ser Ser Leu Val Ala Leu His Gln Ala Val  
85 90 95  
Glu Ala Leu Arg Ser Gly Arg Ile Glu Thr Ala Ile Val Gly Gly Ile  
100 105 110  
Asn Val Ile Ala Ser Pro Ala Ser Phe Ile Ala Phe Ser Gln Ala Ser  
115 120 125  
Met Leu Ser Pro Thr Gly Leu Cys Gln Ala Phe Ser Ala Lys Ala Asp  
130 135 140  
Gly Phe Val Arg Gly Glu Gly Gly Thr Val Phe Val Leu Arg Lys Ala  
145 150 155 160  
Ala His Ala His Gly Ser Arg Asn Pro Val Arg Gly Leu Ile Leu Ala  
165 170 175  
Thr Asp Val Asn Ser Asp Gly Arg Thr Asn Gly Ile Ser Leu Pro Ser  
180 185 190

Ala Glu Ala Gln Glu Val Leu Leu Gln Arg Val Tyr Ser Arg Ala Ser  
 195 200 205

Ile Asp Pro Asn Arg Leu Ala Phe Val Asp Thr His Gly Thr Gly  
 210 215 220

<210> 55

<211> 254

<212> PRT

<213> Organisme Inconnu

<220>

<223> Origine de la séquence :organisme du sol

<400> 55

Pro Gln Gln Arg Val Phe Leu Asp Gly Ile Asp Arg Phe Asp Pro Arg  
 1 5 10 15

His Phe Ala Ile Thr Pro Arg Glu Ala Ile Ser Met Asp Pro Gln Gln  
 20 25 30

Arg Leu Leu Glu Val Thr Trp Glu Ala Leu Glu Arg Ala Gly Val  
 35 40 45

Ala Pro Asp Arg Leu Thr Gly Ser Asp Thr Gly Val Phe Ile Gly Ile  
 50 55 60

Ser Thr Asn Asp Tyr Gly Gln Ile Leu Leu Arg Ala Ser Asp Gln Ile  
 65 70 75 80

Asp Pro Gly Met Tyr Phe Gly Thr Gly Asn Leu Leu Asn Ala Ala Ala  
 85 90 95

Gly Arg Leu Ser Tyr Val Leu Gly Leu Gln Gly Pro Ser Met Ala Val  
 100 105 110

Asp Thr Ala Cys Pro Ser Ser Leu Val Ala Ile His Leu Ala Cys Gln  
 115 120 125

Ser Leu Arg Asn Arg Glu Cys Arg Met Ala Leu Ala Gly Gly Ala Asn  
 130 135 140

Leu Val Leu Val Pro Glu Val Thr Val Asn Cys Cys Arg Ala Lys Met  
 145 150 155 160

Leu Ala Pro Asp Gly Arg Cys Lys Thr Phe Asp Ala Ala Ala Asp Gly  
 165 170 175

Tyr Val Arg Gly Glu Gly Ala Ala Val Ile Val Leu Lys Arg Leu Ser  
 180 185 190  
 Asp Ala Leu Ala Asp Gly Asp Pro Ile Val Ala Leu Ile Arg Gly Ser  
 195 200 205  
 Ala Val Asn Gln Asp Gly Arg Ser Gly Gly Phe Thr Ala Pro Asn Glu  
 210 215 220  
 Leu Ala Gln Gln Ala Val Ile Arg Thr Ala Leu Ala Ala Ala Gly Val  
 225 230 235 240  
 Ala Ala Ser Asp Ile Gly Tyr Val Asp Thr His Gly Thr Gly  
 245 250

&lt;210&gt; 56

&lt;211&gt; 254

&lt;212&gt; PRT

&lt;213&gt; Organisme Inconnu

&lt;220&gt;

&lt;223&gt; Origine de la séquence : organisme du sol

&lt;400&gt; 56

Pro Gln Gln Arg Val Phe Leu Asp Gly Ile Asp Arg Phe Asp Pro Gln  
 1 5 10 15  
 Phe Phe Gly Ile Ala Pro Arg Glu Ala Ala Gly Ile Asp Pro Gln Gln  
 20 25 30  
 Arg Leu Leu Leu Glu Thr Thr Trp Glu Ala Leu Glu Asp Ala Gly Thr  
 35 40 45  
 Ser Pro Glu Lys Leu Gln Gly Thr Pro Ala Gly Val Phe Val Gly Ile  
 50 55 60  
 Asn Ser Ile Asp Tyr Ala Thr Leu Gln Leu Gln Asn Cys Asp Leu Ala  
 65 70 75 80  
 Ser Ile Asp Ala Tyr Ser Leu Ser Gly Ser Ala His Ser Ile Ala Ala  
 85 90 95  
 Gly Arg Leu Ala Tyr Val Leu Gly Leu Gln Gly Pro Ala Met Ala Val  
 100 105 110  
 Asp Thr Ala Cys Ser Ser Ser Leu Val Ala Ile His Leu Ala Cys Gln  
 115 120 125

Ser Leu Arg Asn Asp Asp Cys Arg Val Ala Val Ala Gly Gly Val His  
130 135 140

Val Thr Leu Thr Pro Ile Asn Met Val Val Phe Ser Lys Leu Arg Met  
145 150 155 160

Leu Ala Ala Asp Gly Lys Cys Lys Thr Phe Asp Gly Arg Gly Asp Gly  
165 170 175

Phe Val Glu Gly Glu Gly Cys Ala Val Ile Val Leu Lys Arg Leu Ser  
180 185 190

His Ala Leu Ala Asp Lys Asp Arg Ile Leu Ala Leu Val Arg Gly Ser  
195 200 205

Ala Val Asn Gln Asp Gly Ala Ser Ser Gly Leu Thr Ala Pro Asn Gly  
210 215 220

Pro Ala Gln Glu Ala Val Ile Arg Ala Ala Leu Lys Arg Ala Gly Val  
225 230 235 240

Gln Pro Ala Glu Val Gly Tyr Val Asp Thr His Gly Thr Gly  
245 250

<210> 57

<211> 222

<212> PRT

<213> Organisme Inconnu

<220>

<223> Origine de la séquence :organisme du sol

<400> 57

Pro Gln Glu Arg Val Leu Leu Glu Ser Ser Trp His Ala Leu Glu Asp  
1 5 10 15

Ala Gly Tyr Ala Gly Glu Ser Ile Ala Gly Ala Arg Cys Gly Val Tyr  
20 25 30

Met Gly Phe Asn Gly Gly Asp Tyr Gly Asp Leu Leu Tyr Gly Gln Pro  
35 40 45

Ser Leu Pro Pro His Ala Met Trp Gly Asn Ala Ala Ser Val Leu Ser  
50 55 60

Ala Arg Ile Ala Tyr Tyr Leu Asp Leu Gln Gly Pro Ala Ile Thr Leu  
65 70 75 80

Asp Thr Ala Cys Ser Ser Ser Leu Val Ala Val His Leu Ala Cys Gln  
85 90 95

Gly Leu Trp Thr Gly Glu Thr Asp Leu Ala Leu Ala Gly Gly Val Trp  
100 105 110

Ile Gln Cys Thr Pro Gly Phe Leu Ile Ser Ser Ser Arg Ala Gly Met  
115 120 125

Leu Ser Pro Thr Gly Gln Cys Arg Ala Phe Gly Ala Gly Ala Asp Gly  
130 135 140

Phe Val Pro Ser Glu Gly Val Gly Val Val Val Leu Lys Arg Leu Gln  
145 150 155 160

Asp Ala Leu Asp Ala Gly Asp His Xaa Tyr Gly Val Ile Arg Gly Ser  
165 170 175

Ala Ile Asn Gln Asp Gly Ala Ser Asn Gly Ile Thr Ala Pro Ser Ala  
180 185 190

Ala Ala Gln Glu Arg Leu Gln Arg His Val Tyr Asp Ser Phe Gly Ile  
195 200 205

Asp Ala Ser Arg Leu Gln Met Ile Glu Ala His Gly Thr Gly  
210 215 220

<210> 58

<211> 223

<212> PRT

<213> Organisme Inconnu

<220>

<223> Origine de la séquence : organisme du sol

<400> 58

Pro Gln Glu Arg Val Leu Leu Glu Val Thr Trp Glu Ala Leu Glu Asp  
1 5 10 15

Ala Gly Gln Asp Val Asp Arg Leu Ala Gly Arg Pro Val Gly Val Phe  
20 25 30

Val Gly Ile Ser Ser Asn Asp Tyr Gly Gln Leu Gln Asn Gly Asp Pro  
35 40 45

Ala Asp Val Asp Ala Tyr Val Gly Thr Gly Asn Ala Leu Ser Ile Ala  
50 55 60



Ala Asn Arg Leu Ser Tyr Thr Phe Asp Phe Arg Gly Pro Ser Leu Ala  
65 70 75 80

Val Asp Thr Ala Cys Ser Ser Ser Leu Val Ala Ile His Leu Ala Cys  
85 90 95

Gln Ser Val Arg Arg Gly Glu Ala Glu Leu Ala Val Ala Ala Gly Val  
100 105 110

Asn Leu Ile Leu Thr Pro Gly Leu Thr Val Asn Phe Thr Arg Ala Gly  
115 120 125

Met Met Ala Pro Asp Gly Arg Cys Lys Thr Phe Asp Ala Ala Ala Asn  
130 135 140

Gly Tyr Val Arg Gly Glu Gly Ala Gly Val Val Val Leu Lys Pro Leu  
145 150 155 160

Ala Gln Ala Ile Ala Asp Gly Asp Pro Ile Tyr Ala Ile Val Arg Gly  
165 170 175

Ser Ala Val Asn Gln Asp Gly Arg Ser Asn Gly Leu Thr Ala Pro Asn  
180 185 190

Arg Gln Ala Gln Glu Val Val Leu Arg Ala Ala Tyr Arg Asp Ala Gly  
195 200 205

Ile Ser Pro Ala Asp Val Asp Ala Val Glu Ala His Gly Thr Gly  
210 215 220

<210> 59

<211> 235

<212> PRT

<213> Organisme Inconnu

<220>

<223> Origine de la séquence :organisme du sol

<400> 59

Pro Gln Gln Arg Val Phe Leu Glu Asp Ala Thr Glu Val Asp Val Asp  
1 5 10 15

Ala Leu Ser Asp Gly Glu Asp Val Val Ile Ala Gly Ile Met Gln His  
20 25 30

Ile Glu Glu Ala Gly Ile His Ser Gly Asp Ser Ser Cys Val Leu Pro  
35 40 45

Pro Val Asp Ile Pro Pro Lys Ala Leu Gln Thr Ile Arg Asp His Thr  
 50 55 60  
 Phe Lys Leu Ala Arg Ala Leu Lys Val Ile Gly Leu Met Asn Val Gln  
 65 70 75 80  
 Tyr Ala Ile Gln Arg Asp Lys Val Tyr Val Ile Glu Val Asn Pro Arg  
 85 90 95  
 Ala Ser Arg Thr Val Pro Tyr Val Ser Lys Ala Thr Gly Val Pro Leu  
 100 105 110  
 Ala Lys Val Ala Ser Arg Leu Met Thr Gly Arg Lys Leu His Glu Leu  
 115 120 125  
 Leu Pro Glu Gly Val Glu Arg Gly Trp Ile Thr Thr Ala Gly Glu Asn  
 130 135 140  
 Phe Tyr Val Lys Ser Pro Val Phe Pro Trp Gly Lys Phe Pro Gly Val  
 145 150 155 160  
 Asp Thr Val Leu Gly Pro Glu Met Lys Ser Thr Gly Glu Val Met Gly  
 165 170 175  
 Val Ala Asp Asn Phe Gly Glu Ala Phe Ala Lys Ala Gln Ile Ala Ala  
 180 185 190  
 Gly Thr Tyr Leu Pro Thr Glu Gly Thr Val Phe Ile Ser Val Asn Asp  
 195 200 205  
 Arg Asp Lys Gly Asn Val Ile Gln Leu Ala Gln Arg Phe Ser Glu Leu  
 210 215 220  
 Gly Phe Gly Ile Val Asp Thr His Gly Thr Gly  
 225 230 235

&lt;210&gt; 60

&lt;211&gt; 1269

&lt;212&gt; ADN

&lt;213&gt; Organisme Inconnu

&lt;220&gt;

&lt;223&gt; Origine de la séquence :Organisme du sol

&lt;400&gt; 60

taacaggaag aagcttgctt ctttgctgac gagtggcgga cgggtgagta acacgtggga 60  
 acctgcctta tggctcgga taacgtctgg aaacggagcgc taacaccgga tgtgcccttc 120

```

gggggaaagt ttacgccatg agagggggccc gcgtccgatt aggtagttgg tggggtaatt 180
gcccaccaag ccgacgatcg gtactgtggtc tgagaggatg atcagccaca ctggggactga 240
gacacggccc agactcctac gggaggcgagc agtgggggaat attggacaat gggggcaacc 300
ctgatccagc aatgccgcgt gagtgtgtaa ggccttaggg ttgtaaaagt ctttcgcacg 360
cgacgatgat gacggtagcg tgagaagaag ccccggtcaa ctctgtgcca gcagccgcgg 420
taatacgaag ggggcgagcg ttgttcggaa ttactgggcg taaagggcgc gtatggcgccc 480
cgatcagta gattgtgaag ccccgggctc aacctgggaa ctgcatttga tactgtcggg 540
cttgagtccc ggagaggatg gtggaattcc cagtgtagag gtgaaatcg tagatattgg 600
gaagaacacc ggtggcgaag gcggccatct ggacggacac tgacgctgag gcgcgaaagc 660
gtggggagca aacaggatta gataccctgg tagtccacgc cgtaaacgat gaatgctaga 720
cgctgggggt catgcacttc ggtgtccgag ctaacgcatt aagcattccg cctgggggat 780
acggccgcga ggttaaaact caaagggaatt gacggggggc cgcacacgag ctggagcatg 840
tggtttaatt cgaagcaacg cgcagaacct taccacccct tgacatgtcc attgccggtc 900
cgagagattg gaccttcagt tcgggtggat ggaacacagg tgctgcatgg ctgtcgtcag 960
ctcgtgtcgt gagatgttgg gtttaagtccc gcaacgagcg caaccccctac cgccagttgc 1020
catcattcag ttggggcactc ttgttgaact gccggtgaca agccggagga aggcggggat 1080
gacgtcaagt cctcatggcc cttatgggtt gggctacaca cgtgtctaaa tagcggtgac 1140
agtgggacgc gaagtgcgaa gatggagcaa atccccaaaa gccgtctcag ttccgattgc 1200
actctgaac tcgggtgcat gaagtgggaa tcgctagtaa tcgcggatca gcacgcgcgcg 1260
gtgaatacgc
1269

```

&lt;210&gt; 61

&lt;211&gt; 1500

&lt;212&gt; ADN

&lt;213&gt; Organime Inconnu

&lt;220&gt;

&lt;223&gt; Origine de la séquence : Organisme du sol

&lt;400&gt; 61

```

ttttaaaaag acggccagtg aattgtaata cgactcacta tagggcgcaat tgggcccctct 60
agatgcatacg tcgagcgccc gccagtgtag tgatatctcg cagaattcgc ccttcaggcc 120
taacacatcgc aagtcgaacg agggcttcgg ccctagtggc gcacgggtga gtaaacacgtg 180
ggaacctgcc ttatggttcg ggataacgtc tggaaaacgga cgtaaacacc ggaatgtccc 240
ttcgggggaa agttttacgc atgagagggg cccgcgtccg attagtagtg ttgtggggta 300
atggccccac aagccgacga tcggtagctg gtctgagagg atgatcagcc aactggggac 360
tgagacacgg cccagactcc tacgggaggg agcagtgggg aatattggac aatgggggca 420
acctgatccc agcaatgcgc cgtgagtgat gaaggcctta ggggtgtgaa gctcttcgcg 480
acgcgacgat gatgacggta gcgtgagaag aagccccggc taacttcgtg ccagcagccc 540
cggtaatacag aagggggcga gcgttgttcg gaattactgg gcgtaaaagg ccgtagggcg 600
gcccgatcag tcagatgtga aagccccggg ctcaacctgg gaactgcatt tgatactgtc 660
gggcttgagt tccggagagg atggtggaat tccagtgata gagggtgaaat tcgtagatat 720
tgggaagaac accggtggcg aaggcgccca tctggacgga cactgacgct gaggcgcgaa 780
agcgtgggga gcaaacagga ttatataccc tggtagtcca cgcgtaaac gatgaatgct 840
agacgctggg gtgcattgac ttcggtgtcg ccgctaacgc attaagcatt ccgcctgggg 900
agtacggcgcc caaggttaaa actcaaaagg attgacgggg gccgcgcaca gcggtggagc 960
atgtggttta attcgaagca acgcgcagaa ccttaccac ccttgacatg tccattgccg 1020
gtccgagaga ttggaccttc agttcggctg gatggaaacac aggtgctgca ttgctgtcgt 1080
cagctcgtgt cgtgagatgt tgggttaagt cccgcaacga gcgcaaacccc taccgccagt 1140

```

```

tgccatcatt cagttgggca ctctggtgga actgccggtg acaagccgga ggaaggcggg 1200
gatgacgtca agtccctcatg gcccttatgg gttggggtac acacgtgcta caatggcggt 1260
gacagtggga cgcgaagtgc caagatggag caaatcccca aaagccgtct cagttcggtat 1320
tgacatctgc aactcgggtg catgaagtgt gaatcgctag taatcgcgga tcagcacgcc 1380
gggtgaata cgttcccggt ccttgtacac accgcccag ggcaattcc agcacactgg 1440
cggccgttac tagtggatcc gagctcggta ccaagcttgg cgtaatcatg gtcatagtctg 1500

```

<210> 62  
 <211> 1366  
 <212> ADN  
 <213> Organisme Inconnu

<220>  
 <223> Origine de la séquence :Organisme du sol

```

<400> 62
acgacggcca gtgaattgta atacgactca ctatagggcg aattggggccc tctagatgca 60
tgctcgagcg gccgccagtg tgatggatat ctgcagaatt cgcccttcag gcctaacaca 120
tgcaagtcga acgaaggctt cggccttagt ggccgcacggg tgagtaaacac gtgggaacct 180
gcctttcggt tcggaataaac gtctggaaac ggacgctaac accgatacgc cccttcgggg 240
gaaagttcac gccgagagag gggcccgctg cggattaggt agttggtgag gtaatggctc 300
accaagcctt cgatccgtag ctggtctgag aggatgatca gccacactgg gactgagaca 360
cggcccagac tcctacggga ggacgacgtg gggaattatt gacaatgggc gcaagcctga 420
tccagcaatg ccgcgtgagt gatgaaggcc ttagggttgt aaagctcttt cgcacgcgac 480
gatgatgacg gtacgctgag aagaagcccc ggctaacttc gtgccagcag ccgcggtaat 540
acgaaggggg ctacgcttgt tcggaattac tggcgtaaaa gggcgcgtag gcggcctgct 600
tagtcagaag tgaaaacccc gggctcaacc tgggaatagc ttttgatact ggcaggcttg 660
agttccggag aggatggttg aattccagtg tagaggtga aattcgtaga tattgggaag 720
aacaccggtg gcgaaggcgg ccatctggac ggacactgac gctgaggcgc gaaagcgttg 780
ggagcaaaaca ggattagata ccctggtagt ccacgcgcta aacgatgaat gctagacgtc 840
ggggtgcagt cactcgggtg tcgcccgtaa ccattaagc attccgcctg gggagtacgg 900
ccgcaaggtt aaaactcaaa ggaattgacg ggggcccgca caagcggttg agcatgtggt 960
ttaattcgaa gcaacgcgca gaaccttacc aaccttgac atgtccatta tgggcttcag 1020
agatgaggtc ctctcagttcg gctgggttga acacaggtgc tgcatggctg tcgtcagctc 1080
gtgtcgtgag atgttgggtt aagtcgccga acgagcgcaa cccctaccgt cagttgccat 1140
cattcagttg ggcactcttg tggaaacgcg ggtgacaagc cggaggaagg cgggatgacg 1200
gtcaagtcct catggccctt atgggttggg ctacacacgt gctacaatgg cgggtgacgt 1260
gggaagcgaa gtcgcgagat ggagcaaatc cccaaaagcc gtctcagttc ggtatgcact 1320
ctgcaactcg agtcgctgaa gttggaatcg ctagtatacg cggatc 1366

```

<210> 63  
 <211> 1360  
 <212> ADN  
 <213> Organisme Inconnu

<220>  
 <223> Origine de la séquence :Organisme du sol

&lt;400&gt; 63

```

acagctatgtg ccatgattac gccaaagcttg gtaccgagct cggatccact agtaacggcc 60
gccagtgtagc tggaaattcgc ccttcaggcc taacacatgc aagtcgaacg ccccgcaagg 120
ggagtgagcag acgggtgagt aacgcgtggg aacataccct ttctctcgga atagctccgg 180
gaaactggaa ttaataccgc atacgccta cgggggaaaag atttatcggg gaaggattgg 240
cccgcgttgg attagctagt tgggtggggta aaggcctacc aaggcgacga tccatagctg 300
gtctgagagg atgatcagcc acattgggac tgagacacgg cccaaactcc tacgggaggc 360
agcagtgagg aatattggac aatggggcga agcctgatcc agccatcgcg cgtgagtgat 420
gaaggcctta ggggtgtaaa gctctttcac cggagaagat aatgacggta tccggagaag 480
aagccccggc taacttcgtg ccagcagcgg cggtaatacg aaggggggcta gtgtgtgtcg 540
gaattactgg cgttaaaagg cactagggcg gatatttaag tcaggggtga aatcccagag 600
ctcaactctg gaactgcctt tgatactggg tatcttgagt atggaagagg taagtggaa 660
tccgagtgtg gaggtgaaat tcgtagatat tcggaggaaac accagtgccg aaggcgccct 720
actggcccat tactgacgct gaggtgcgaa agcgtgggga gcaaacagga ttagataccc 780
tggtagtcca cgcgttaaac gatgaatgtt agcgtcgagg cagtatactt ttccgtggcg 840
cagctaaacg attaaacatt ccgcctgggg agtaacggtc caagatataa actcaaaagg 900
attgacgggg gcccgacaaa gcggtggagc atgtggttta attcgaagca acgcccagaa 960
ccttaccagg tcttgacatt cgggggtttg gcagtgagga cattgtcctt cagttaggct 1020
ggccccagaa cagggtctgc atggctgtcg tcagctcgcg tcgtgatag ttgggttaag 1080
tcccgcacag agcgcacacc tcgcccctag ttgccagcat ttagttaggc actctaaagg 1140
gactgcggcg gataaacgga gaggaagggt gggacgagct caagtcctca ttggccctac 1200
gggctgggct acacacgtgc tacaatgggt gtgacagtg gcagcgagac agcgatgtcg 1260
agctaacttc caaaagccat ctcagttcgg attgcaactc gcaactcgag tgcattgaag 1320
tggaatcgct agtaatcgca gatcagcatg tgcggtgaat 1360

```

&lt;210&gt; 64

&lt;211&gt; 1288

&lt;212&gt; ADN

&lt;213&gt; Organisme Inconnu

&lt;220&gt;

&lt;223&gt; Origine de la séquence :Organisme du sol

&lt;400&gt; 64

```

tccaggaaac agctatgacc atgattacgc caagcttggt accgagctcg gatccactag 60
taacgcgcgc cagttgtctg gaattcgccc ttcaggccta acacatgcaa gtcgacgcgc 120
ccgcaagggg agcggcagac ggggtagtaa cgcgtgggaa tctaccatc cctacggaac 180
aactccggga aactggagct aataccgtat acgcccttgg ggggaaagat ttatcgggga 240
tggtatggcc cgcgttggat tagctagtgt gtggggtaaa ggcctaccaa ggcgacgac 300
catagctggt ctgagaggat gatcagccac attgggactg agacacggcc caaactccta 360
cgggaggcag cagtggggaa tattggacaa tggggcgaag cctgatccag ccattgccgc 420
gtgagtgtag aaggctcttag gattgtaaag ctctttcac ggagaagata atgacgggtat 480
ccggagaaga agccccgctt aactttcgtg ccagcagcgg cggtaatacg aagggggcta 540
gcgttgttgc gaattactgg gcgtaaagcg cactgaggcg gatattttaa tcagggggtga 600
aatcccagag ctcaactctg gaactgcctt tgatactggg tatcttgagt atggaagagg 660
taagtggaaat tgcgagtgtg gaggtgaaat tcgtagatat tcgcaggaaac accagtgggc 720
aaggcgccct actggtccat tactgacgct gaggtgcgaa agcgtgggga gcaaacagga 780
ttagataccc tggtagtcca cgcgttaaac gatgaatgtt agcgttcggc aagtttactt 840
gtcgggtggc cagctaacgc attaaacatt ccgcctgggg agtacggtcg caagattaaa 900

```

```

actcaaagga attgacgggg gcccgacaa gcggtggagc atgtggttta attcgaagca 960
acgcgcagaa ccttaccagc ccttgacatg cccggacagc tacagagatg tagtgttccc 1020
ttcggggacc gggacacagg tgctgcatgg ctgtcgtcag ctctgtctgt gagatgttgg 1080
gttaagtccc gcaacgagcg caaccctcgc ccttagttgc cagcattcag ttgggcactc 1140
taaggggact gccggtgata agccgagagg aagtggggat gacgtcaagt cctnatggcc 1200
cttacgggct gggctacaca cgtctacaa tgggtgggtga cagtggggcag cgaaggaaac 1260
atcccgagct aatctccaaa agccatct 1288

```

&lt;210&gt; 65

&lt;211&gt; 1386

&lt;212&gt; ADN

&lt;213&gt; Organisme Inconnu

&lt;220&gt;

&lt;223&gt; Origine de la séquence :Organisme du sol

&lt;400&gt; 65

```

cgacggccag tgaattgtta tacgactcac tataggcgga attgggccct ctatagtcac 60
gctcgagcgg cgcagctgt gatggatc tgcagaattc gcccttcagg cctaacacat 120
gcaagtcgag cgggcgtagc aatacgtcag cggcagacgg gtgagtaacg cgtgggaaca 180
taccttttgg ttcggaaaca cacagggaac cttgtgctaa taccggataa gcccttacgg 240
ggaaagattt atcgccgaaa gattggcccg cgtctgatta gctagtgtgt agggtaatgg 300
cctaccacag cgacgatcag tagctggctc gagaggatga tcagccacat tgggactgag 360
acacggccca aactcctacg ggaggcagca gtggggaata ttggacaatg ggcgcaagcc 420
tgatccagcc atgcgcgctg agtgatgaag gccctagggt tgtaaagctc ttttgtgcgg 480
gaagataatg acggtaccgc aagaataaag cccggctaac ttctgtccag cagcccggtt 540
aatacgaagg gggctagcgt tgctcggaaat cactgggcgt aaagggtgcg taggcgggtc 600
tttaagtcag ggggtgaatc ctggagctca actccagaac tgctcttgat actgaagatc 660
ttgagttcgg gagaggtgag tggaaactgcg agtgtagagg tgaattctgt agatattcgc 720
aagaacacca gtgggcgaag gcggctcact ggcccagata tgacgctgag gcacgaaagc 780
gtggggagca aacaggatta gataccctcg tagtccacgc cgtaaacgat caatgccagc 840
cgtagtgagg ttactcact agtggcgcat ctaacgcttt aagcattccg cctggggagt 900
acggtcgcaa gattaaaaat caaaggaatt gacgggggccc cgcacaagcg gtggagcatg 960
tggttttaatt cgacgcaacg cgcagaacct taccagccct tgacatgtcc agaccgggtc 1020
gcagagatgt gaccttctct tcggagcctg gagcacaggt gctgcatggc tctcgtcagc 1080
tcgtgtcgtg agatgttggg ttaagtcocg caacgagcgc aacccccctc cttagtgtct 1140
accattttagt tgagcactct aaggagactc ccggtgataa gccgcgagga aggtggggat 1200
gacgtcaagt cctcatggcc cttacgggct gggctacaca cgtctacaaa tggcggtgac 1260
aatgggacgc aggggggcaa cccttcgcaa atctcaaaaa gccctgtctca gttcggattg 1320
ggcttcgcaa ctcgagccca tgaagttgga atcgtagta atcgtggatc agcacgccac 1380
ggtgaa 1386

```

&lt;210&gt; 66

&lt;211&gt; 1223

&lt;212&gt; ADN

&lt;213&gt; Organisme Inconnu

&lt;220&gt;

<223> Origine de la séquence :Organisme du sol

<400> 66

```

agcgccagag ggtgagtaac gcgtgggaat ctacccatct ctacggaaca actccgggaa 60
actggagcta ataccgtata cgtccttcgg gagaaagatt tatcggagat ggaatgagccc 120
gcgttgagta agctagtctg tggggtaaat gcctaccaag gcgacgatcc atagctgggtc 180
tgagaggatg atcagccaca ctgggactga gacacggccc agactcctac gggaggcagc 240
agtggggaat attggacaat gggcgaaagc cggatccagc catgccgcgt gagtgatgaa 300
ggccctaggg ttgtaaagct ctttcaacgg tgaggataat gacggtaacc gtagaagaag 360
ccccggctaa cttcgtgcca gcagcccgcg taatacgaag ggggctagcg ttgttcggaa 420
ttactggggc taaagcgcac gtaggcgagc tattaagtca ggggtgaaat cccgggggtc 480
aaccgccgaa ctgcctttga tactggtagt ctgagtcgc gaagaggtga gtggaattcc 540
gagtgtagag gtgaaattcg tagatattcg gaggaacacc agtggcgaag gcggctcact 600
ggtcgggtac tgacgctgag gtgcgaaagc gtggggagca aacaggatta gataccctgg 660
tagtccacgc cgtaaacgat ggaagctagc cgttggcaag tttacttgtc ggtggcgagc 720
ctaaccgatt aagcttccc cctggggagt acggtcgcaa gattaaaact caaagggaatt 780
gacggggggc cgcacaagcg gtggagcatg tggtttaatt cgaagcaacg cgcagaacct 840
taccagccct tgacatcccg gtcgcgggta ccagagatgg tatccttcag ttcggtcgga 900
ccggtgacag gtgctgcag gctgtcgtca gctcgtgtcg tgagatggtg ggttaagtcc 960
cgcaacgagc gcaaccctcg cccttagttg ccagcattca gttgggcact ctaaggggac 1020
tgccggtgat aagccgagag gaaggtgggg atgacgtcaa gtcctcatgg cccttacggg 1080
ctgggtctaca cagctgctac aatggtgggt acagtgggca gcgagaccgc gaggctcgagc 1140
taatctccaa aagccatctc agttcggatt gcactctgca actcagatgc atgaagtgtg 1200
aatcgctagt aatcgggagt cag 1223

```

<210> 67

<211> 1237

<212> ADN

<213> Organime Inconnu

<220>

<223> Origine de la séquence :Organisme du sol

<400> 67

```

cccgagggg agtggcagag ggtgagtaac gcgtgggaat ctaccccttt ctacggaaca 60
actgagggaa acttcaagcta ataccgtata cggccgagag gcgaaagatt tatcggagaa 120
ggatgagccc gcgttggatt agctagtctg tggggtaaaag gcctaccaag gcgacgatcc 180
atagctggtc tgagagggat atcagccaca ctgggaactga gacacggccc agactcctac 240
gggaggcagc agtggggaaat attggacaat gggcgcaagc ctgattccagc catgccgcgt 300
gagtgatgaa ggccttaggg ttgtaaagct ctttcacgg tgaagataat gacggtaacc 360
ggagaagaag ccccggttaa cttcgtgcca gcagcccgcg taatacgaag ggggctagcg 420
ttgttcggat ttactggggc taaagcgcac gtaggcgagc tattaagtca ggggtgaaat 480
cccggggtcc aaccgccgaa ctgcctttga tactggtagt cttgagttcg aaagaggtga 540
gtggaattcc gagtgtagag gtgaaattcg tagatattcg gaggaacacc agtggcgaag 600
gcggctcact ggctcgatac tgacgctgag gtgcgaaagc gtggggagca aacaggatta 660
gataccctgg tagtccacgc cgtaaactat gagagctagg cgtcgggagc tatactgttc 720
ggtgggcagc ctaacgcatt aagctcttcg cctggggagt acggtcgcaa gattaaaact 780
caaagggaat gacggggggc cgcacaagcg gtggagcatg tggtttaatt cgaagcaacg 840
cgcagaacct taccagccct tgacatcccg atcgcggtta ccagagatgg tatccttcag 900

```

```

ttaggctgga tcggtgacag gtgctgcatg gctgtcgtca gctcgtgtcg tgagatgttg 960
ggttaagtcc cgcaacgagc gcaacccctcg cccttagttg ccatcattca gttggggcact 1020
ctaaggggac tgccggtgat aagccgagag gaaggtgggg atgacgtcaa gtctctatgg 1080
cccttacggg ctgggctaca cacgtgtcac aatgggtggc acagtgggca gcgagaccgc 1140
gaggtcgagc taatctccaa aagccatctc agttcggatt gcactctgca actcagatgc 1200
atgaagtgtg aatcgctagt aatcgtggat cagaatg 1237

```

&lt;210&gt; 68

&lt;211&gt; 1346

&lt;212&gt; ADN

&lt;213&gt; Organisme Inconnu

&lt;220&gt;

&lt;223&gt; Origine de la séquence :Organisme du sol

&lt;400&gt; 68

```

acgacggggc agtgaattgt aatacagctc actatagggc gaattgggccc ctctagatgc 60
atgctcgagc ggccgccagt gtgatggata tctgcagaat tcgcccttca ggccaaacac 120
atgcaagtgc aacggatccc ttcggattag tggcggacgg gtgagtaaca cgcgggaacg 180
tgccctttgg ttcggaacaa ctacgggaaa cttgagctaa taccggataa gcccttccag 240
ggaaagattt atcgccattg gagcgcccg cgtaggatta gctagtgtgt gaggtaaaag 300
ctcaccaagg cgacgatcct tagctgggtc gagaggatga tcagccacat tgggactgag 360
acacggccca aactcctacg ggaggcagca gtggggaatc ttgcgcaatg ggcgaaagcc 420
tgacgcagcc atgcccgctg aatgatgaag gtcttaggat tgtaaaaatc ttccaccggg 480
gacgataatg acggtaccgc gagaagaagc cccggctaac ttcgtgccag cagcccggtg 540
aatacgaagg gggctagcgt tgctcggaat tactgggcgt aaagggagcg taggcggata 600
gtttagtcag aggtgaaagc ccagggtcta accttggaaat tgcccttgat actggctatc 660
ttgagtacgg aagaggtatg tgsaactccg agtgtagagg tgaatttcgt agatattcgg 720
aagaacacca gtggcgaaag cgacatactg gtcggttact gacgctgagg ctgcgaaagc 780
tggggagcaa acaggattag ataccctggt agtcccgcgt gtaaacgagt agtgctagtt 840
gtcggcatgc atgcatgtcg gtggcgcgag taacgcatta agcactccgc ctggggagta 900
cggtcgcaag attaaaaact aaaggaattg acggggggccc gcacaagcgg tggagcatgt 960
ggtttaatcc gaagcaacgc gcagaacctt accacctttt gacatgcccg gaccgtccca 1020
gagatggagc ttcccttcg gggactggga cacaggtgct gcatggctgt cgtcagctcg 1080
tgtcgtgaga tgttgggtta agtcccgcga cgagcgcaac cctcgctatt agttgccatc 1140
aggtttggtt gggcactcta ataggaccgc cgggtggtaag ccggaggaag gtggggatga 1200
cgtaagctcc tcatggccct tacaaggtgg gctacacag tgctacaatg gcgactacag 1260
agggtgcaa tcccgcgagg gggagccaat ccctaaagat cgtctcagtt cggattgcac 1320
tctgcaactc gagtgcatag agtttg 1346

```

&lt;210&gt; 69

&lt;211&gt; 1500

&lt;212&gt; ADN

&lt;213&gt; Organisme Inconnu

&lt;220&gt;

&lt;223&gt; Origine de la séquence :Organisme du sol



&lt;400&gt; 69

acagctatga ccatgattac gccaaagcttg gtaccgagct cggatccact agtaacggcc 60  
 gccagtgtgc tggaaatcgc ccttcaggcc taacacatgc aagtcgaacy ccgtagcaat 120  
 acggagtgcc agacgggtga gtaaacacgtg ggaacgtgcc ctttggttcg gaacaacaca 180  
 gggaaacttg tgctaatacc gaataagccc ttacggggaa agatttatcg ccaaggatc 240  
 ggcgccgctg tgattagcta gttggtgggg taacggccca ccaaggctac gatcagtagc 300  
 tggctctaga ggatgatcag ccacactggg actgagacac ggcccagact cctacgggag 360  
 gcagcagtta ggaatcttgg acaatgggcg caagcctgat ccagccatgc cgcgtgagt 420  
 atgaaggcct taggggtgta aagctcttcc agcggggaag ataatgacgg taccgcgaga 480  
 agaagccccc gctaaactcg tgccagcagc cgcggttaata cgaaggggcg tagcgttgct 540  
 cggaatcact gggcgtaaaag cgacgttagg cggatcttta agtcaggggg gaaactctgg 600  
 agctcaactc cagaactgcc ttgtactg gggatctcga gtccggaaga ggtgagtga 660  
 actccagtg tagaggtgaa attcgtagat attcgaaga acaccagtgg cgaaggcgcc 720  
 tcactggctc ggtactgacg ctgaggtgcg aaagcgtggg gagcaaacag gattagatac 780  
 cctggtagtc cagcgctgaa acgatggatg ctgcccgttg gcggggttac tcgtcagtg 840  
 cgacgtaaac gcattaaaga tcccgcctgg ggagtacggt cgcaagatta aaactcaagg 900  
 gaattgacgg gggcccgca aagcgggtgga gcatgtggtt caattcgaag caacgcgcag 960  
 aaccttacc a gcccttgaca tgtcccgtat ggacttcaga gatgaggtcc ttcaagtctgg 1020  
 ctggggggaa cacagggtgct gcatggctgt cgtcagctcg tgtcgtgaga tgttggggtta 1080  
 agtcccccaa cgagcgcaac cctgcgccct agttgccatc atttagttgg gcactctaag 1140  
 gggactcggc gtgataagc gcgaggaagg tggggatgac gtcaagtct catggccctt 1200  
 acgggctggc ctacacacgt gctacaatgg cggtgacagt gggagcgaat ggagcaatcc 1260  
 tgcgcaaatc tcaaaaagcc gtctcagttc ggattggggg ctgcaactcg accccatgaa 1320  
 gtcggaatcg ctgtaatcg cagatcagca cgctgcggtg aatacgttcc cgggccttgt 1380  
 acacaccgcc caagggcgaa ttctgcagat atccatcaca ctggcgccg ctcgagcatg 1440  
 catctagagg gcccaattcg ccctatagtg agtcgtatta caattcactg gccgtcggtt 1500

&lt;210&gt; 70

&lt;211&gt; 1113

&lt;212&gt; ADN

&lt;213&gt; Organisme Inconnu

&lt;220&gt;

&lt;223&gt; Origine de la séquence :Organisme du sol

&lt;400&gt; 70

gagctaatac cgtataatga cttcgggtcca aagatttacc gcctgaggat gagcccgct 60  
 cggattagct agttggtagg gtaaaaagcct accaaggcga cgtatcgtag ctggtctgag 120  
 aggatgatca gccacactgg gactgagaca cggccagac tcctacggga cgcagcagtg 180  
 gggaaatttg gacaatgggc gcaagcctga tccagcaatg ccgcgtgagt gatgaaggcc 240  
 ttagggttgt aaagctcttt taccggggaa gataatgact gtaccgggag aataagcccc 300  
 ggctaaactc gtgccagcag ccgcggtaat acggaggggg ctacgcttgt tcggaattac 360  
 tggcgctaaa gcgcacgtag cgcgctttgt aagttagagg tgaagcccg gggtcactac 420  
 ccggaattgc ctttaagact gcatcgctcg aattgtggag aggttaagtgg aattccgagt 480  
 gttagagtgaa aattcgtaga tattcggaag aacaccagtg gcgaaggcga cttactggac 540  
 acatatcgac gctgaggtgc gaaagcgtgg ggagcaaaa ggattagata cctcggtagt 600  
 ccacgcgcta aacgatgatg actagctgtc ggggcgctta cgttttcggt ggcgacgcta 660  
 acgcgttaag tcattccgct ggggagtagc gccgcaaggt taaactcaaa gaaattgacg 720  
 ggggcctgca caagcgggtg agcatgtggt ttaattcgaa gcaacgcgca gaaccttacc 780

```

agcgtttgac atgccaggac ggtttccaga gatggattcc ttcccttacg ggacctggac 840
acagggtcgtg catggctgtc gtcagctcgt gtcgtgagat gttggggttaa gtccccgaac 900
gagcgcgaacc ctgcgtcttta gttgctacca tttagttgag cactctagag aaactgcgg 960
tgataagcgc gaggaagggtg gggatgacgt caagtctca tggcccttac gcgctgggct 1020
acacacgtgc tacaatggcg gtgacaacgg gcagcaaac gcgcgagagt agcaaatccc 1080
gaaaagccgt ctcagttcgg attgttctct gca 1113

```

&lt;210&gt; 71

&lt;211&gt; 1225

&lt;212&gt; ADN

&lt;213&gt; Organisme Inconnu

&lt;220&gt;

&lt;223&gt; Origine de la séquence :Organisme du sol

&lt;400&gt; 71

```

ggagcggcgg acgggtgagt aacgcgtggg aacgtgccct ttggtacgga acaactgagg 60
gaaacttcag ctaataaccgt atgtgccctt cgggggaaag atttatcgcc attggagcgg 120
cccgcgttgg attaggtagt tgggtgggta aaggcctacc aagcctacga tccatagctg 180
gtctgagagg atgatcagcc acactgggac tgagacacgg ccagactcc tacgggagcg 240
agcagtaggg aatcttgcgc aatggggcaa agcctgacgc agccatgcgg cgtgtatgat 300
gaaggcttta ggattgtaa atactttcac cggggaagat aatgacggta cccggagaag 360
aagccccggc taactctcgt ccagcagccg cggtaatcag aaggggggta gcgtgtctcg 420
gaattactgg gcgtaaaagg cgcgtaggcg gatatttaag tcgggggtga aagcccaggg 480
ctcaaccctg gaattgcctt cgatactgga tatcttgagt tcgggagagg tgagtggaat 540
gcgagtgta gaggtgaaat tcgtagatat tcggcggaac accagtggcg aaggcgactc 600
actggcccca tactgacgct gaggcgcgaa agcgtgggga gcaaacagga ttatagatccc 660
tggtagtcca cgcgtgaaac gatgagtgtc agttgtcggc atgcatgcat tcggtgacg 720
cagctaaccg attaagcact ccgcctgggg agtacggtcg caagattaaa actcaaagga 780
attgacgggg gcccgacaaa gcgggtggagc atgtggttta attcgaagca acgcgcagaa 840
ccttaccacc tttgacatg cctgcatcgc tggagagatc cagttttccc ttcggggaca 900
gggacacagg tgctgcatgg ctgtcgtcag ctgcgtgtcg gagatgttgg gttaatgtccc 960
gcaacgagcg caaccctcgc cattagtgc catcattaag ttgggcactc taatgggacc 1020
gcgggtggta agcgggagga aggtggggat gacgtcaagt cctcatggcc cttacggggg 1080
gggctacaca cgtgctacaa tggcgactac agaggggttc aaacctgcga aggggagcta 1140
atccctaaaa gtgcgtctcag ttcggattgc actctgcaac tcgagtgcat gaagtcggaa 1200
tcgctagtaa tcgcgatca gcatg 1225

```

&lt;210&gt; 72

&lt;211&gt; 1286

&lt;212&gt; ADN

&lt;213&gt; Organisme Inconnu

&lt;220&gt;

&lt;223&gt; Origine de la séquence :Organisme du sol

&lt;400&gt; 72

```

atgattagta gcaataactaa tcgatgacga gcggcggacg ggtgagtaat acgtaggaa 60

```

```

ctgcccttaa gcgggggata actaaggga actttagcta ataccgcata aactcgagag 120
agaaaagctg cagcaatgtg gcacttgagg aggggcctgc gtcagattag ctagtgtgtg 180
aggtaataagc tcaccaaggc gatgatctgt aactggctcg agaggacgac cagtcacact 240
gggactgaga cacggccag actcctacgg gaggcagcag tggggaatat tggacaatgg 300
gggcaacccg gatccagcga tgccgcgtgg gtgaagaagg ccttcggggt gtaaaagcct 360
ttaggctggg aagaaggtta gttagaggaa tgctattaac ttgacggtag cgacagaata 420
agcacggcga aactctgtgc cagcagccgc ggtaatacag aggggtcgag cgttaactcg 480
atttactggg cgtaaaaggg gcgtaggcgg tgagatgtgt gtgatgtgaa agccccagcg 540
tcaacctggg aagtgcatcg caaactgtct gactggagta tatgagaggg tggcggaatt 600
tcgggtgtag cgggtgaaatg cgtagagatc ggaaggaaac tcgatggcga aggcagccac 660
ctggcataat actgacgctg aggcgcgaaa gcgtggggat cgaacaggat tagataccct 720
ggtagtccac gctgtaaact atgagtaact gatgttggtg ggggaacct tcggtagtca 780
agctaacggc ataagtattc gcctcgggaa gtacggccgc aagggtgaaa ctcaaatgaa 840
ttgacggggg ccgcgacaa cgggtggagca tgtggtttaa ttcgatgcaa cgcaagaac 900
cttacctacc ctgacatcc tgagaatctg gcttagtagc tggagtgccg aaaggagctc 960
agagacagg gctgcattgc tgcgtcagc tgcgttgtg agatgttggg ttaagtcccg 1020
taacgagcgc aaccttgcc cttagttgcc atcatttagt tggggactct aaggggaccg 1080
ccagtgatga actggaggaa ggcggggacg acgtcaagtc atcatggcct ttatgggtag 1140
ggccacacac gtgctacaat ggggcgtacg gagggtcgca aaccgcgag ggggagctaa 1200
tctcataaag cgtctcgtag tccggattgg agtctgcaac tcgactccat gaagttggaa 1260
tcgctagtaa tcgcgaatca gccattg 1286

```

&lt;210&gt; 73

&lt;211&gt; 1288

&lt;212&gt; ADN

&lt;213&gt; Organisme Inconnu

&lt;220&gt;

&lt;223&gt; Origine de la séquence :Organisme du sol

&lt;400&gt; 73

```

cggggcaacc ctggcgcgga gcggcgaaac ggtgagtaat gcatcggaac gtgtctctct 60
gtgggggata accagtcgaa agactggcta ataccgcatg agatcgaaag atgaaagcag 120
gggacgcgcaa ggccttgccg gagaggagca gccgatcccg gattagctag ttggtggggg 180
aaaagcctac caaggcgacg atccgtagct ggtctgagag gacgaccagc cacactggga 240
ctgagacatc gccagactc ctacgggagg cagcagtggg gaattttgga cagtggggcg 300
aacctcgatc cagccatgcc gcgtgtgtga agaaggcctt cgggttgtaa agcactttcg 360
gacggaaacga aatcgcgcga gttaatagtt cgcgtggatg acggtaccgt aagaagaagc 420
accgctcaac tacgtgccag cagccgcggt aatacgtagg gtgcgagcgt taatcggaat 480
taactggcgt aaagtgtgcg caggcggcct cgcaagtcga gtgtgaaatc cccgagctta 540
acttgggaat tgcgctcgaa actacggagc cggagtgtgg cagaggaagg tggaaattcca 600
cgtgtagcgt tgaaatgcgt agagatgttg aggaacaccc atggcgaaag cggcctttcg 660
ggccaaacact gacgctcatg caccgaaagcg tggggagcaa acaggattag ataccctggt 720
agttccacgc ctaaaacgat atgactagtt gttggaggag ttaaatcctt tagtaacgca 780
gctaacgcgt gaagtcatcc gcctggggag tacggtcgca agattaaaac tcaaaaggaat 840
tgacgggggg ccgcacaagc ggtggatgat gtggtttaat tcgatgcaac gcgaaaaaac 900
ttacactacc ttgacatgct aggaacgctg cgggtgcgcc aggggaaacc aaagggaacc 960
tagacacagg tgctgcatgg ctgctgctcg ctgctgtcgt gagatgttgg gttaaagtc 1020
gcaacgagcg caaacctcgc cattagtgc tacattagtt tgagcactct aatgggactg 1080

```

```

ccggtgacaa accggaggaa ggtggggatg acgtcaagtc ctcatggccc ttatgggtag 1140
ggctacacac gtcatacaat ggcgcgtaga gaggggtgccc aacccgcgag ggggagccaa 1200
tcccagaaag cgcgtcgtag tccggattgg agtctgcaac tcgactccca tgaagtcgga 1260
atcgtagta atcgcggatc agcatgtc 1288

```

<210> 74  
 <211> 600  
 <212> ADN  
 <213> Organisme Inconnu

<220>  
 <223> Origine de la séquence :Organisme du sol

```

<400> 74
cgtgccagca gccgcggttaa tacgtaggtg gcaagcgttg tccggaatta ttgggcgtaa 60
agcgcgcgca ggtggtttct taagtctgat gtgaaagccc acggcttaac cgtggagggt 120
cattggaac tgggagactt gagtgcagaa gaggaaagtg gaattccaag tgtagcggtg 180
aaatgcgtag agatttggag gaacaccagt ggcgaaggcg actttctggt ctgcaactga 240
cgctgaggcg cgaagcatg gggagcaaac aggattagat accctggtag tccatgccgt 300
aaacgatgag tgctaagtgt taggggggttt ccgcccctta gtgctgcagc taacgcatta 360
agcactccgc ctggggagta gcaccgcaag gttgaaactc aaaggaattg acggggggccc 420
gcacaagcgg tggagcatgt ggtttaattc gaagcaacgc gaagaacctt accaggtctt 480
gacatccgat tganocgtct agagatagag ttttcccttc ggggacattg gtgacaggtg 540
gtgcatgggt gtgcgtcagct cgtgtcgtga gatgttgggt taagtccgcg aacgagcgca 600

```

<210> 75  
 <211> 601  
 <212> ADN  
 <213> Organisme Inconnu

<220>  
 <223> Origine de la séquence :Organisme du sol

```

<400> 75
cgtgccagca gccgcggttaa tacgtaggtg gcaagcgttg tccggaatta ttgggcgtaa 60
agcgcgcgca ggtggtttct taagtctgat gtgaaagccc acggcttaac cgtggagggt 120
cattggaac tgggagactt gagtgcagaa gaggaaagtg gaattccaag tgtagcggtg 180
aaatgcgtag agatttggag gaacaccagt ggcgaaggcg actttctggt ctgcaactga 240
cgctgaggcg cgaagcatg gggagcaaac aggattagat accctggtag tccatgccgt 300
aaacgatgag tgctaagtgt taggggggttt ccgcccctta gtgctgagct aacgcattaa 360
gcactccgc tggggagtaac gaccgcaagg ttgaaactca aaggaattga cgggggccc 420
cacaagcggg ggagcatgtg gtttaattcg aagcaacgcg aagaacctta ccaggtcttg 480
acatccgat gacgctctag agatagagtt ttccctcgg ggaacattgtg cacaggtggt 540
gcatggttgt cgtcagctcg tgctcgtgaga tgttgggtta agtcccgcga cgagcgacc 600
c 601

```

<210> 76

&lt;211&gt; 1236

&lt;212&gt; ADN

&lt;213&gt; Organisme Inconnu

&lt;220&gt;

&lt;223&gt; Origine de la séquence :Organisme du sol

&lt;400&gt; 76

```

tgccttctgtag acgggggataa cttcgggaaa cgggagctaa tacccggataa tcctctctccc 60
cacatggggga agagttgaaa ggcgctttcg cgtcactaca ggatggggccc cgggtgcatt 120
agctagtgtgg tagggtaacg gcctaccaag gcgacgatgc atagccgacc tgagaggggtg 180
atcgggccaca ttgggactga gacacggccc aaactcctac gggaggcgagc agtaggggaat 240
cttccacaat ggacgaaagt ctgatggagc aacgcccgtg gagtgatgaa ggttttcgga 300
tcgtaaaact ctgttgaag ggaagaacca gtacgtcagg caatggacgt accttgacgg 360
taccttatta gaaagccacg gctaactacg tgccagcagc cgcggtaata cgtaggtggc 420
aagcgttgtc cggaaattatt gggcgtaaa ggcgcgcagg tggtttctta agtctgatgt 480
gaaagcccac ggccttaaccg tggaggggtca ttggaaactg ggagacttga gtgcagaaga 540
ggaaagtgga attccaagt tagcggcgaa atgcgtagag atttgaggga acaccagtg 600
cgaaggcgac tttctgtct gcaactgacg ctgaggcgcg aaagcatggg gagcaaacag 660
gattagatac cctggtagtc catgctgtaa acgatgagtg ctaagtgtta ggggggtttcc 720
gcccttagt gctgcagcta acgcattaag cactccgctc ggggagtagc accgcaaggt 780
tgaataccta aggaattgac gggggccgcg acaagcggtg gagcatgtgg ttaattcga 840
agcaacgcga agaaccctac caggctctga catcccgatg atcgctctg agatagagtt 900
ttcccttcgg ggacattggt gacaggtggt gcattggtgt cgtcagctcg tgcgtgaga 960
gtttgggta agtcccgcaa cgagcgcaac ccttaattct agttgccatc atttagttgt 1020
gcactctaag gtgactgccg gtgataaac ggaggaaggt ggggagtagc tcaaatcatc 1080
atgccctta tgacctggcg tacacactg ctacaatgga cggtaacaa agtcgcgtaac 1140
tcgcgagagt atgctaact catagaaccg ttctcagttc ggattgtagg ctgcaactcg 1200
cctacatgaa gccggaaatc ctagtaatcg cggatc 1236

```

&lt;210&gt; 77

&lt;211&gt; 815

&lt;212&gt; ADN

&lt;213&gt; Organisme Inconnu

&lt;220&gt;

&lt;223&gt; Origine de la séquence :Organisme du sol

&lt;400&gt; 77

```

caagcgttgt ccggaattat tgggcgtaaa gagctcgtag gcggtttgtc gcgtctgctg 60
tgaaaaactcg aggcctcaacc tcgggcttgc agtgggtacg gccagactag agtcgggtag 120
gggtgactgg aattcctggt gtacggctgg aatgcgcaga tatcaggagg aacaccgatg 180
gcgaaggcag gtcactgggc cgcaactgac gctgaggagc gaaagcatgg ggagcgaaca 240
ggattagata ccttggtagt ccatgccgta aacgttgggc actaggtgtg gggctcattc 300
cacgagttcc gtgcgcgagc aaacgcatta agtgcgccgc ctggggagta cggccgaag 360
gcttaaaact caaagaattt gacgggggcc gcacaaagcg gcggagcatg cggattaatt 420
cgatgcaacg cgaagaacct taccaaggct tgacatacac cggaaacttc cacagatggt 480
tgccccgcaa ggtcggtgta caggtggtgc atggttgtcg tcagctcgtg tcgtgaagat 540
gttggggttaa gtcccgcaac gagcgcaacc ctgcctcat gttgccagca cgtgatgggt 600

```

```

gggactcata ggagactgcc ggggtcaact cggaggaagg tggggatgac gtcaaatcat 660
catgcccctt atgtcttggg ctacacgcat gctacaatgg ccggtacaaa gggctgcgat 720
accgcaagg ggagcgaatc ccaaaaagcc ggtctcagtt cggattgggg tctgcaactc 780
gacccatga agtcggagtc gctagtaatc gcaga 815

```

&lt;210&gt; 78

&lt;211&gt; 826

&lt;212&gt; ADN

&lt;213&gt; Organisme Inconnu

&lt;220&gt;

&lt;223&gt; Origine de la séquence :Organisme du sol

&lt;400&gt; 78

```

tcgtagggtg cttgtcacgt cgggtgtgaa agcttggggc ttaactccag gtctgcattc 60
gatacgggct ggctagaggt aggtaggggg gaacggaatt cctggtgtag cggtgaaatg 120
gcagatatac aggaggaaca ccggtggcga agcggtttct ctgggacctt cctgacgctg 180
aggagcgaaa cgttggggag cgaacaggat tagataccct ggtagtccac gctgtaaaacg 240
ttgggcgcta ggtgtgggga ccttccacgg ttccgcgccc gtactaacg cattaagcgc 300
cccgctctgg gactacggcc gcaaggctaa aactcaaaagg aattgacggg ggcccgacac 360
agcggcggaag catgttgcct aattcgacgc aacgcgaaga acctaccac ggcttgacat 420
gcgccggaaa gcttcagaga tggagccctc ttccgactgg gtgacaggtg gtgcatggct 480
ctcgtcagct cgtgtcgtga gatgttgggt taagtcccg ccaacgagcga accttgttc 540
aatgttgcca gcaaatcctc tcgggttggt tggggaacta ttggaagctg ccgggggtcaa 600
ctcggaggaa ggtggggacg acgtcaagtc atcatgcccc ttatgtcttg ggtgcaaac 660
atgtacaaat ggcgggtaca gagggttgcg ataccgcaag gtggagcgaa tccctaaaag 720
ccggtctcag ttccgattgg ggtctgcaac tcgaccccat gaagtccgag tcgctagtaa 780
tcgcagatca gcaacgctgc ggtgaatacg ttcccgggcc ttgtac 826

```

&lt;210&gt; 79

&lt;211&gt; 799

&lt;212&gt; ADN

&lt;213&gt; Organisme Inconnu

&lt;220&gt;

&lt;223&gt; Origine de la séquence :Organisme du sol

&lt;400&gt; 79

```

cgtaggcggt ttgtcgcgtc tgccgtgaaa gtccggggct caactccgga tctgcggtgg 60
gtacggggag actagagtga tgtaggggag actggaattc ctggtgtagc ggtgaaatgc 120
gcagatatac ggaggaacac cgtaggcgaa ggcaggtctc tgggcattaa ctgacgctga 180
ggagcgaaa catggggagc gaacaggatt agataccctg gtactccatg ccgtaaacgt 240
tgggcactag gtgtggggga cattccacgt ttccgcgccc gtactaacg cattaagtgc 300
cccgctctgg gactacggcc gcaaggctaa aactcaaaagg aattgacggg ggcccgacac 360
agcggcggaag catcgggatt aattcgatgc aacgcgaaga acctaccac ggcttgacat 420
gaaccggaaa caactgaaa cagggtgccc gcttgcgctc ggtttacagg tgggtcatgg 480
ttgtcgtcag ctcgtgctgt gagatgttgg gtttaagccc gcaacgagcg caacctcgtg 540
tctatgttgc cagcgcgtta tggcggggac tcataggaga ctgcccgggt caactcggag 600

```

```

gaaggtgggg acgacgtcaa atcatcatgc cccttatgtc ttgggcttca cgcagtctac 660
aatggcgggt acaaaaggggt gcgatactgt gaggtggagc taatcccaaa aagcgggtct 720
cagttcggat tggggctctgc aactcgaccc catgaagtgc gagtgcgtag taatcgaca 780
tcagcaacgc tgcggtgaa 799

```

&lt;210&gt; 80

&lt;211&gt; 1250

&lt;212&gt; ADN

&lt;213&gt; Organime Inconnu

&lt;220&gt;

&lt;223&gt; Origine de la séquence :Organisme du sol

&lt;400&gt; 80

```

tgccagctgt ctggtggatt agtggcgaac gggtagtaaa cactgagta acctgccctt 60
aactctggga taagctgggg aaactgggtc taatgcgga tatgactcct catcgcatgg 120
tggggggggg aaagcttttt gtgggttttg atggactcgc ggccatcag ctgtgttggtg 180
aggtaatggc tcaccaaggc gacgacgggt agcggcctg agaggggtgac cgccacact 240
gggactgaga cagcgccag acttctacgc gaggcagcag tggggaatat tgcacaatgg 300
gcgaagacgt gatgcagca cgcgcgtga gggatgacgg cctcgggtt gtaaacctct 360
ttcagtaggg aagaagcgaa agtgacggta cctgcagaag aagcgcggc taactacgtg 420
ccagcagccg cggtaatacg tagggcgcaa gcgttatccg gaattattgg gcgtaaaag 480
ctcgtagggc gtttgtcgc tctgccgtga aagtcgggg ctcaactccg gatctgcgt 540
gggtacgggc agactagagt gatgtagggg agactggaat tcctggtgta gcggtgaaat 600
gcgcagatat caggaggaa acgatgggcg aaggcaggtc tctgggcat aactgacgt 660
gaggagcgaa agcatgggga gcgaacagga ttagataccc tggtagtcca tgcgtaaac 720
gttgggcact aggtgtgggg gacattccac gttttccgc cgttagctaa cgcattaagt 780
gccccgcctg gggagtacgg ccgcaaggct aaaactcaaa ggaattgacg ggggcccgca 840
caagcggcgg agcatgcgga ttaattcgat gcaacgcgag gaaccttacc aaggcttgac 900
atgaaccgga aatacctgga aacagggtgc ccgcttgctg tcggtttaca ggtggtgcat 960
ggtgcgcgtc agctcgtgtc gtgagatgtt gggttaaagt ccgcaacgag gcgaaccttc 1020
gttctatgtt gccagcgcgt tatggcgggg actcatagga gactgcgggg gtcaactcgg 1080
aggaaggtgg ggacgacgtc aaatcatcat gcccttatg tcttgggctt caagcatgct 1140
acaatggcgc gtacaaaggg ttgcgatact gtgaggtgga gctaactcca aaaagccgg 1200
ctcagttcgg attggggtct gcaactcgac cccatgaagt cggagtgcgt 1250

```

&lt;210&gt; 81

&lt;211&gt; 1210

&lt;212&gt; ADN

&lt;213&gt; Organime Inconnu

&lt;220&gt;

&lt;223&gt; Origine de la séquence :Organisme du sol

&lt;400&gt; 81

```

cgtaataacc ggatacggcg cgagagtctt cggactttcg cgagaaagat tcgcaaggat 60
cactgagggg cgagcctgcg gcccatcagc tagttggtga ggtaagagct caccaaggct 120
aagacgggta gctgggtcga gaggatgac agccacactg gaactgagac acggtccaga 180

```

```

ctctcactcggg aggcagcagct gggggaatatt gcgcaatggg cgaaagcctg acgcagccac 240
gccgcgtgag cgatgaggggc cttcgggtcg taaagctctg tggggagaga cgaataaagg 300
cgggtgaagag tcggccttga cgggtatctcc tttagcaagca ccggctaact ccgtgccagc 360
agcccgcggtg atacggagggg tgcaaacctgt gctcggaaac attggggcgta aagcgcacgt 420
agggcgcgctg ataagttggg tgtgaaagcc ctggggtcaa cccagggaagt gcattcaaaa 480
ctgtcacgtg tgaatctcgg aggggggtcag agaattcccg gtgtagaggt gaaattcgta 540
gatactcgga ggaataccag tggcgaaggc gctggcctgg acgaagattg acgctgaggt 600
gcgaaagcgc ggggagcaaa caggattaga taccctggta gtcgcgcgtg taaacgatga 660
gtgctagacg ggggagggtat tgaccccttc gctgccgaag ctaacgcggt aagcactccg 720
cctggggagt acggtcgcaa gactaaaact caaaggaatt gacgggggcc cgacacaagg 780
gtggagcatg tggtttaatt cgacgcaacg ccaaaacct tacctgggtt aaatccgcg 840
gaacctggct gaaaggctgg ggtgccctcc ggggaatcgg tgagaagggt ctgcatggct 900
gtcgtcagct cgtgtcgtga gatgttgggt taagtcccgc aacgagcgca acccctatcg 960
tcagttgccca acattaaggt gggaaactct gcgagactgc cgtctctaac cggaggaagg 1020
tggggacgac gtcaagtctc catggccctt atgccagggg ctacacacgt gctacaatgg 1080
ctggtaacaat gagccgcmaa accgcgaggt caagctaat ctaaaaaaac agtctcagtt 1140
cggatcgagg tctgcaactc gactcogtga agctggaatc gctagtaatc gaagatcagc 1200
acgctttcgg 1210

```

&lt;210&gt; 82

&lt;211&gt; 1272

&lt;212&gt; ADN

&lt;213&gt; Organisme Inconnu

&lt;220&gt;

&lt;223&gt; Origine de la séquence :Organisme du sol

&lt;400&gt; 82

```

gatgccagct tgctggtgga ttagtggcga acgggtgagt aacacgtgag taacctgccc 60
ttaactctgg gataagcctg ggaactcggg tctaataccg gatatgactc ctcatcgcat 120
gggtgggggtt ggaagctttt ttgtggtttt ggatggactc gcggcctatc agctgtgttg 180
tgaggtaatg gctcaccaag gcgacgacgg gttagccggc tgagagggtg accggccaca 240
ctggggactga gacacggccc agactcctac gggaggcgag agtgggggaat attgcacaat 300
gggcgaaagc ctgatgcagc gacgcgcgct gagggatgac ggccttcggg ttgtaaacct 360
ctttcagtag ggaagaagcg aaagtgcagg tacctgcaga agaagcgccg gctaactacg 420
tgccagcagc cgcggttaata cgtaggggcg aagcgttatc cggaattatt gggcgtaaaag 480
agctcgtagg cggtttgtcg cgtctgccgt gaaagtccgg ggtcacaactc cggatctcgc 540
gtgggtacgg gcagactaga gtgatgtag ggagactgga attcctgggt tagcggtgaa 600
atgcgcagat atcaggagga acaccgatgg cgaaggcgag tctctgggca ttaactgacg 660
ctgagggaac aaagcatggg gagcgaacag gattagatac cctggtgatc catgccgtaa 720
acgttgggca ctagggtgtg gggacattcc acgttttccg ccgcgtagct aacgcattaa 780
gtgccccgc tggggagtag gccgcgaagg ctaaaaactca aaggaattga cggggggccc 840
cacaagcgcg ggagcatgcg gattaattcg atgcaacgcg aagaacctta ccaaggcttg 900
acatgaacgg gaaataacct gaaacagggt cccgccttgc ggtcggttta caggtgtgtgc 960
atggtgtgtg tcagctcgtg tctgtgagat ttgggttaag tccgcgaacg agcgcacaacc 1020
tcgtttctatg ttgccagcgc gttatggcgg ggactcatag gagactcgcg ggtgcacaact 1080
ggaggaaggt ggggacgacg tcaaatcatc atgccctta tgccttgggc ctacgcgatg 1140
ctacaatggc cggtaacaaag ggttgcgata ctgtgaggtg gagctgatcc caaaaagccg 1200
gtccacgattc ggattggggg ctgcaactcg accccatgaa gtccgagtcg ctagttaatc 1260

```



cagatcagca ac

1272

&lt;210&gt; 83

&lt;211&gt; 1247

&lt;212&gt; ADN

&lt;213&gt; Organisme Inconnu

&lt;220&gt;

&lt;223&gt; Origine de la séquence :Organisme du sol

&lt;400&gt; 83

```

tgtttagtag caataactaaa tgaatgacgag cggcgggacgg gtgaggaaca cgtagggaacc 60
tgcccaagag aggggggacaa ccaagggaataa ctttggtctaa taccgcataa tctctacgga 120
gaaaagttag ccgtaagggt ggcgcttttg gaggggcctg cgtccgatta gttagtgtgt 180
gaggttaatag ctaccaaga ctgtgatcgg taactggtct gagaggacga ccagtcacac 240
tgggactgag acacggccca gactcctacg ggaggcagca gtggggaatc ttggacaatg 300
ggggcaacc tgatccagcg atgccgcgtg ggtgaagaag gccttcgggt tgtaaagccc 360
tttaggcggg gaagaaggat atgggatgaa taagcctgta ttttgacggt acccgagaa 420
taagcaccgg caaactctgt gccagcagcc ggcgtaatc agagggtgag agcgttaac 480
ggatttactg ggcgtaaaag ggcgtaggc ggttggtgta gtgtgatgtg aaagcccg 540
gctcaacctg ggaagtgcac cgcaaacgac acaactggag tatatgagag ggtggcgaa 600
tttcgggtgt agcgtgaaa tgcgtagaga tcggaaggaa cgtcgatggc gaaggcagcc 660
acctggcata atactggcg tgaggcgca aagcgtgggg agcgaaacag attagatacc 720
ctggtagtca cggcgtgaaa cgatgagaac tagatgttgg agggggaacc ctccagatc 780
gaagctaacg cgataagttc tccgcctggg aagtaacgac gcaagactga aactcaaaag 840
aattgacggg gggccgcaca agcgggtggg catgtggttt aattcgatgc aacgcgaaga 900
accttacctg cccttgacat cctgcgaatc ttgccgagag gtgagatgag cgcagggagc 960
gcagagacag gtgctgcacg gctgtcgta gctcgtgttg tgagatgttg ggttaagtcc 1020
cgtaacgagc gcaacccctt tccttagttg ccatcattta gttggggact ctaaggagac 1080
cgccgggtgat gaaccggagg aaggcgggga cgacgtcaag tcatcatggc ctttatgggt 1140
agggctacac acgtgctaca atggggcgta cagaggggtc ccaacccgcg agggggagcc 1200
aatctcttaa agcgtctcgt agtcgggatt ggaagtctgca actcgac 1247

```

&lt;210&gt; 84

&lt;211&gt; 1292

&lt;212&gt; ADN

&lt;213&gt; Organisme Inconnu

&lt;220&gt;

&lt;223&gt; Origine de la séquence :Organisme du sol

&lt;400&gt; 84

```

ggctcgcaag agcaaccggc gaacgggtgc gtaacacgtg aacaacctgc cctcgtgtg 60
gggatagccg ggctaaccgc cgggtaatac cgcatacgtt ctctctgggg agtctctggg 120
agaggaaagc tccggcgacg ggggaggggt tcgcgcccta tcagctagtt gggcggttaa 180
tgccccacca aggcgacgac gggtagctgg tctgagagga tggccagcca cattgggact 240
gagagacggc ccagactcct acgggaggca gcagtgggga atcttgcga atggccgaaa 300
ggctgacgca ggcagccgcg gtgtgggagg acgccttctg ggtgtgaaac cactgttgcc 360

```

```

cgggacgaac agcctcttct gagaggtctg acggtaccgg gtgaggaagc accggctaac 420
tccgtgccag cagccgcggt aatacggagg gtgcgagcgt tgtccggaat catctgggcgt 480
aaagggcgcg taggtggccc ggtcagttcg tggtagaaagc gcggggctca accctgcgtc 540
ggccatgaat actgccgcgg ctggagcact gtgagggcag gcggaattcc ggggtgtacg 600
gtggaatggt tagagatcgc gaagaacacc ggtggcgaag gcggcctgct gggcagtagc 660
tgacactgag gcgcgacagc gtggggagca aacaggatga gatccctgg tagtccacgc 720
cgtaaacgat gggcactagg cgcttggggg agcgaccccc cgaggggcgg cgctaaccga 780
ttaagtgcgc cgcttgggga gtacggccgc aaggctgaaa ctcaaggaa ttgacggggg 840
cccgcacaag cgggtggagc tgtggtttaa ttgcagcga cgcaagaac ctacctagg 900
cttgacatac acgggaaacc ggtcagaacc ggcgggcctt ctccggagcc cgtgcacagg 960
tgctgcatgg ctgctgtcag ctgctgtcgt gagatgttg gttaagtcgc gcaacgagcg 1020
caacccctgt ctctagtgtc cagcgcgtca tggcggggac tctagagaga ctcccggtgc 1080
caaaccggag gaaggtgggg atgacgtcaa gtcactatg tcttaagctc tagggctaca 1140
cacgtgtctac aatggcgggg acagaggggc gcgagccagg aacggcaagc caatcccgtc 1200
aaccocgctc cagttcggat tgtcgtctgc aactcgacgg catgaagctg gaatcgctag 1260
taatcgtgga tcagctacgc caccggtaac ac

```

&lt;210&gt; 85

&lt;211&gt; 1300

&lt;212&gt; ADN

&lt;213&gt; Organisme Inconnu

&lt;220&gt;

&lt;223&gt; Origine de la séquence :Organisme du sol

&lt;400&gt; 85

```

tcccttcggg agcaagtaca gcggcgaaac ggtgagtaac acgtaggtaa cctaccctgg 60
agactgggat aacctgcgga aaggcgggct aataccagat aagaccacga gggctgcggc 120
ccttggggca aaaggtggcc tctacttgta agctaccact ccgggatggg cctgcgcgcc 180
attagctagt tggcggggta acggccaccc aaggcagaga tggctagctg gtctgagagg 240
atggccagcc acacagggac tgagacacgg ccagactcc tacgggaggg agcagtgagg 300
aatattgcgc aatgggcgaa agcctgacgc agcgacggcg cgtgggtgat gaagcccttc 360
gggtcgtaaa gcctgtcaa gagggacgaa accttgtcga cctaacacgt cggcaacctg 420
acggtaacct tgaaggaaac accggctaac tccgtgccag cagccgcggt aatacggagg 480
gtgcgagcgt tgttcggaat tactggcgct aaagcgcgtg tagggcgccct ctctagctgt 540
gtgtgaaagc ccggggctca accccggaag tgcattggt actgggaggg ttgagtacgg 600
gagaggaggg tggaaattcct ggtgtagcgg tgaatgcgt agatactagg aggaacacct 660
gtggcggaag cggccctctg gacggatact gacgtgaga cgcgaaagcg tggggagcaa 720
acaggattag atacctgggt agtccacgct gtaaacgatg ggcactaggt gttcggggta 780
ttgaccctct gagtgcgcga gctaacgcat taagtgcgcc gcctggggaa tacggccgca 840
aggttaaaac tcaaaaggaat tgacgggggg ccgcacaagc ggtggagcat gtggtttaat 900
tcgacgaac gcgaagaacc ttacctgggc tagacaacct cagaatagag 960
gtctccccgc aaggggcggg tggttcaggt gctgcatggc tgtcgtcagc tctgtcgtgt 1020
agatgtttgg ttaagtcccg caacgagcgc aaccctgtc tctagtgtct accatctagt 1080
tgagcactct agagagactg ccnngtgtta aacggggagg aggtggggag gacgtcaagt 1140
cctcatggcc cttatgtcca gggctacaca cgtgctacaa tgggcgatac aaaggggtgc 1200
gaaccgcgga ggggaagcca atcccaaaaa gtcgctctca gttcggattg gagctgcgaa 1260
ctcgactcca tgaaggcgga atcgctagta atcgcggatc

```

<210> 86  
 <211> 1186  
 <212> ADN  
 <213> Organisme Inconnu

<220>  
 <223> Origine de la séquence :Organisme du sol

<400> 86  
 caatgggcag cggcggacgg gtgagtaaca cgtgggaatg tacctttcgg tgcggaacaa 60  
 ctacagggaac cttgagctaa tgccgcatac gcccttacgg ggaagattt atcgccgaaa 120  
 gatcagcccg cgttgatta gctagtgggt gaggtaatgg ccaccaaagg cgacgatcca 180  
 tagctggttt gagagaacga ccagcctcac tgggactgag acacggccca gactcctacg 240  
 ggaggcagca gttgggaatc ttggacaatg ggggaaaccc tgatccagcc atgcccggtg 300  
 agtgatgaag gccttcgggt tgtaaaactc tttcgacggg gacgataatg acggtaccgg 360  
 tagaagaagc tccggctaac ttcgtgccag cagccgcggt aatacgaagg gggctagcgt 420  
 tgttcggaa tactgggctg aaagcgtgcg caggcggcta tccaagtcag tggtgaaagc 480  
 ccggagctca actccggaat tgccattgaa actgtttagc ttgagtacga gagaggtgag 540  
 tggaataccc agtgtagagg tgaatctcgt agatattggg tagaacaccg gtggcggaagg 600  
 cggctcactg gctcgtaact gacgctcagg cagcagacg tggggatcaa acaggattag 660  
 atacccttgg agtccacgcc gtaaacgatg aacgctagcc gtgggatagc ttgctattca 720  
 gtggcgcgag taacgcatta agcgttccgc ctggggagta cgcccgcaag gttgagactc 780  
 agaggaattg acggggggccc gcacaagcgg tggagcatgt ggtttaattc gacgcaacgc 840  
 gcagaacctt accagggttt gacatcctgt gctcgccggt gaaagccggt tttcccgcaa 900  
 gggacgcaga gacaggtgct gcatggctgt cgtcagctcg tgcgtgaga tgttgggtta 960  
 agtccccgaa cgagcgcaac cctcgcttt agttgccatc attcagttgg gcacttaga 1020  
 gggaccgccc ggcacaagcc ggagggaagg ggggatgacg tcaagtcccc atggccctta 1080  
 caccctggcg tacacacgtg ctacaatggc ggtgacagtg ggacagagct cgcgagagtc 1140  
 agctaataccc aaaaaaccgt cccagttcag attgcactct gcaact 1186

<210> 87  
 <211> 1454  
 <212> ADN  
 <213> Organisme Inconnu

<220>  
 <223> Origine de la séquence :Organisme du sol

<400> 87  
 cgacggccag tgaattgtaa tacgactcac tataggcgca attgggacct ctatagtcac 60  
 gctcgagccg ccgccagtgat gatggatc tgacagaattc gcccttcagg cctaacacat 120  
 gcaagctcag cgagaaaggc cgcttcggcg cctgagtaca gcggcgacag ggtgcgtaac 180  
 acgtgggcaa tctgtccttg agatggggat aaccacagca agtttggct aataccgaat 240  
 aagactacag gaggcaactc ccgtggttaa aggggtgctct ctgcggggag catgcgcttg 300  
 aggaggagcc cgcggcctat cagctagtgt gtagggtcac ggccctacaa ggcgaagacg 360  
 ggtagctggt ctgagaggat gaccagccac acggggactg agacacggcc ccgactccta 420  
 cgggaggcag cagtggggaa tattgggcaa tgggggaaac cctgaccacg cgacgcgcg 480  
 tgggtgatga aggccttcgg tcgtgaaagc cctgtcgggc ggaacgaagg ttctcacggc 540

```

aaatagccgt gagaggtgac ggtaccgccc aaggaagcac cggccaactc cgtgccagca 600
gccgcggttaa gacggagggt gcaagcgctg ctccgaatca ctgggcgttaa agggctcgta 660
ggcggtctcg caagtctggc gtgaaagccc aaggctcagc ctctggaagt cgctcgaaac 720
tgccgagctg gagtgccgga ggggagagtg gaattccccg tgtagcggtg aaatgcgtag 780
agatcgggag gaataccggt ggcgaaagcg actctctgga cggcaactga cgtgtaggca 840
cgaaagcgtg gggagcaaac aggattagat accctggtag tccacgcgcg aaacgatgga 900
cactaggtgt cgggggtatc cactcccccg gtgcccgcgc taacgcagta agtgtcccg 960
ctgggaagta cggtcgcaag attaaaactc aaagggaattg acggggggccc gcacaagcgg 1020
tgaggacatgt ggttcaatc gatgcaacgc gaagaacctt acctggggtt gacatctggc 1080
gaatctctgg gaaaccagag agtgcccgcg ggggagcgcc aagacaggtg ctgcatggct 1140
gtcgtcagct cgtgccgtga ggtgttgggt taagtccccg aacgcagcga acccttacc 1200
ttagttgccc ccgggtcaag ccgtggcact ccaagggaaac tgcccgtgtt aagcgggagg 1260
aagggtggga cgacgtcaag tcatcatggc ctttatatcc aggggtacac acgtgctaca 1320
atggctggga canagcgtg ccaacgcgcg agcgggagct aatcgcaaaa ccccgacctc 1380
agttcggatc ggagctctga actcgactcc gtgaagctgg aatcgctagt aatcgcgat 1440
cagcatgcgc cggt

```

<210> 88  
 <211> 1307  
 <212> ADN  
 <213> Organime Inconnu

<220>  
 <223> Origine de la séquence :Organisme du sol

```

<400> 88
cccttcgggg agcgagtaca gcgcgcaacg ggtgagtaac acgtaggtaa cctaccctgg 60
tgactgggat aacttgccga aaggcgggct aataccagat aagaccacga gggctgcggc 120
ctttggggta aaagatggcc tctgcttgca tgcatacagc ccgggatggg cctgcgcgcc 180
attagctagt tggtaggta acggctcacc aaggcagaga tggttagctg gtctgagagg 240
atggccagcc acctgggac tgagacacgc ccagactcc tacgggaggg acagtgggg 300
aatattgcgc aatgggagaa agcctgacgc agcgacgcgc cgtgggtgat gaaggccttc 360
gggtcgtaaa gcctgtcaa gagggacgaa acctcgccga cccaatacgt cggcgacctg 420
acggtaacct tgaaggaaag accggttaac tccgtgccag cagccgcggt aatacggagg 480
gtgcaagcgt tgttcggaat cactgggcgt aaagcgcgtg tagggcgccct tcttagtctg 540
gtgtgaaagc cgggggtcca accccggaag agcattggat actggaaggc tggagtagcc 600
gagaggaggg tggaaattcct ggtgtagcgg tgaatgcgt agatatcagg aggaacacgc 660
gtggcggaagg cggccctctg gacggatact gacgtgaga cgcgacagcg tggggagcaa 720
acaggattag atacctggt agtcacgcgc gtaaacgatg ggtactagggt gttcggggta 780
ttgacccctc gagtgcgcga gctaaccgat taagtaccgc gcctggggac tacggccgca 840
aggetaaaa tcaaaagaaat tgacgggggc ccgcacaagc ggtggagcat gtggtttaat 900
tgacgcaac gcgaagaaacc ttacctgggc tagacaaacac tggacagccc cagaaatggg 960
gtcttccccg aagggaactg tggttcaggt gctgcattggc tctctgcagc tctgtctgtg 1020
agatgttctg ttaagtcgag caacgagcgc aacctctgct tctagttgct accataagt 1080
tgagcaactc agagagactg ccggtgttaa acgggaggaa ggtggggagc agctcaagtc 1140
ctcatggccc ttatgtccag ggctacacac gtgctacaat ggacagtaca aagggtctgc 1200
aaccctgtag ggggagccaa tccccaaaag ctgttctcag ttccgattgg agtctgcaac 1260
tcgactccat gaaggcggaa tcgctagtaa tcgcgtagca gcatgcc 1307

```

<210> 89  
 <211> 1305  
 <212> ADN  
 <213> Organisme Inconnu

<220>  
 <223> Origine de la séquence :Organisme du sol

<400> 89  
 gggagcaatc cccaagtaga gcggcgaaac ggtgagtaac gcgtgggttaa tctgcctccg 60  
 agtgggggaa aacatcgga aactggtgct aataccgcat aacatcggtg ggtcttcgga 120  
 tctgacgata aaagccgggg accgcaaggc ctggcgcttg gagaggagcc gcgctccgat 180  
 tagctagtgt gtggggtaata ggcccaccaa ggcttcgata ggtagccggc ctgagagggc 240  
 ggacggccac actgggactg agacacggcc cagactccta cgggaggcag cagtggggaa 300  
 tttttcgcaa tggcgaaag cctgacgaag caacgcgcgc tggaggatga gggccttcgg 360  
 gtctaaact cctgtcgacc gggacgaaag taggatggcc taatacgccg atctattgac 420  
 tgtaccggtg gaggaagcca cggctaactc tgtgccagca gccgcggtaa tacagagggtg 480  
 gcaagcgttg ttcggaatta ctgggcgtaa agggcgcgta ggcggcttgg tcagtcccg 540  
 gtgaaatccc tcggctcaac tgaggaaactg cacgggaaac tgcttggtt gagttcggga 600  
 gaggggaagt gtattccggg ttagcggtg aaatgcgtag atatccggag gaacaccggt 660  
 ggcgaaggcg gcttctctga ccgacactga cgctaggcg cgaaagctag gggagcaaac 720  
 gggattagat accccggtg tcttagctgt aaacgatgag tgctgggtgt aggggggtatc 780  
 aacccccct gtgcggaagc taacgcatta agcaactcgc ctggggagta cggtcgcaag 840  
 gctgaaactc aaaggaattg acggggggccc gcacaagcgg tggagcatgt ggttcaattc 900  
 gacgcaacgc gaagaacctt accgggggtt gaactgtac ggacagctct agagatagag 960  
 tcttctcttg ggaccctgac agaggtgtgt catggctgtc gtcagctcgt gtcgtgagat 1020  
 gttgggttaa gtcccgaac gagcgcaacc cttgcctcct gttgccatca ggtaaagctg 1080  
 ggcactctgg agagactgcc ggtgataaac cggaggaagg tggggatgac gtcaagtctt 1140  
 catggccttt atgccccggg ctacacacgt gctacaatgg ccggtacaaa gggtcgcaaa 1200  
 accgagaggt ggagctaata ccaaaaagcc ggtcccagtt cggattgcag tctgcaactc 1260  
 gactgcatag agttggaatc gctagtaatc gcggatcagc atgcc 1305

<210> 90  
 <211> 1299  
 <212> ADN  
 <213> Organisme Inconnu

<220>  
 <223> Origine de la séquence :Organisme du sol

<400> 90  
 gggctttcgg gtcctgagta aagtggcgaa cgggtgagta acgcgtaggt aacctgacct 60  
 cgagtgtgga ataacctggc gaaagccggg ctaataccgc atgacgtctt cgggtcttcg 120  
 gacttgagga ccaaaagggtg cgagctttga gcctgtcgc tcgagaaggg gctcgtcc 180  
 cattagctag ttgggtgggt gatggcctac caaggcgacg atgggtagcc gggctgagag 240  
 gctgtccgg cacaactgaa ccgagacacg gtccagactc ctacgggagg cagcagtggtg 300  
 gaatcttgcc caatggggga aacctgacg caacgacgcc gcgtgggcga tgaaggcctt 360  
 cgggtcgtaa agccctgtcg agcgggacga accgtgcgag ctataacata gctcgtgctc 420

```

gacgggtaccg cttagaggaag ccccggttaa ctcctgtgcca gcagcgcgag taatacggag 480
ggggctagcg ttattcggaat ttattgggag taaaggcgct gtaggcggtc ctgtgtgtcc 540
catgtgaaag ccccggtc caccgggaa ctgcatggga aactgcggag cttgagtcg 600
ggagagtgta gtggaattcc cagtgtagcg gtgaaatgag tagatattgg gaggaacacc 660
agtgggcgaag gcggctcact ggaccggtag tgacgtgag acgcgaaggc cagggggagca 720
aacgggatta gatacccggg tagtcctggc tgtaaacgat gagcacttgg ttggcggggt 780
atcgaccctt gccgtgtgta agctaaccga ttaagtgtcc cgcctgggga gtacggccgc 840
aaggctgaaa ctcaaaggaa ttgacggggg cccgcacaag cgggtggagca tgtgggtcaa 900
ttcgacgcaa cgcgaagaac cttacctggg tttgaactgc aggtgacagc cctgaaaagg 960
gggtcttctt tcgggacacc tgtagaggtg ccgcatggct gtcgtcagct cgtgtcgtga 1020
gatgttgggt taagtccgcg aacgagcgca accctactc ctagtgcga cgggtctggc 1080
cgggaactct agggggaccg ccggtgataa accggaggaa ggtgggggat acgtcaagtc 1140
ctcatggcct ttatgtccag ggctacacac gtgctacaac ggacggtaca aagggtctgcg 1200
aaggcgcgag ccggagccaa tcccaaaaag ccgttctcca gtgcggtatg cagtctgcga 1260
ctcgactgca tgaaggtgga atcgctagta atcgcggt 1299

```

&lt;210&gt; 91

&lt;211&gt; 1296

&lt;212&gt; ADN

&lt;213&gt; Organisme Inconnu

&lt;220&gt;

&lt;223&gt; Origine de la séquence :Organisme du sol

&lt;400&gt; 91

```

atgtctggta gcaataccag atgatggcaa gtggcgagcg ggtgagtaat acgtagggat 60
ctgcccgaa gagggggaca acccggggaa actcgggcta ataccgcata ctattctgag 120
gaagaaaagt tggcgcaagc caggcgcttt tggaggaacc tacgtccgat tagctagtgt 180
gtgaggtaaa ggctcaccaa ggcagagatc ggtagctggt ctgagaggat gaccagccac 240
actgggactg agacacggcc cagactccta cgggagcgag cagtggggaa tattggacaa 300
tggggcgcaac cctgatccag cgatgccgag tgtgtgaaga aggccttcgg gttgtaaaag 360
actttagttg gggaaagaat aatgtttttt aatagagagc attgttgacg gtaccacaaag 420
aataagcacc ggctaactct gtgccagcag ccgcggtaat acagaggggt caagcggtaa 480
tcggagttac tggcggttaa gggcgcgtag gcggtgttgc aagtgagatg tgaaatccct 540
gggtctaacct taggaaccgc attttagact gcaatgctag agtacagtag agggtagttg 600
aatttccggt gtacgggtga aatgcgtaga gatcgggaag aacaccagtg gcgaaggcca 660
ctacctggac tgacactgac gctgaggcgc gagagcgtgg ggagcaaaaca ggattagata 720
ccctggtagt ccacgctgta aacgatgaga actagatgtt ggtgcgcgag agccgcaaac 780
tatcgaagct aacgcgataa gttctccgac tggggagtag gccgcgaagg ttaaaactca 840
aaggaaatga cgggggcccgc cacaagcggt ggagcatgtg gtttaattcg atgcaacgcg 900
aggaacctta cctacccttg acatccacag aatttgatag agatatcgaa gtgcccgaag 960
gaactgtgag acaggtgtcg ctggctgtgc gtcagctcgt gttgtgagat gttgggttaa 1020
gtcccgtaac gagcgcacac cttactctta gttgccaa caataatggt gggactctaa 1080
ggagactgcc ggtgaagaac cggaggagag tggggacgac gtcaagtcac catggccttt 1140
atgggtaggg ctacacacgt gctacaatgg ggcgtacaga ggggtgcgca cgtgcgaagg 1200
ggagccaatc ccggaaaagc cctcgtatgc cagattgaag tctgcaactc gacttcatga 1260
agtgcgaatc gctagtaatc gcgaatcaga acgttc 1296

```

<210> 92  
 <211> 1250  
 <212> ADN  
 <213> Organisme Inconnu  
 <220>  
 <223> Origine de la séquence :Organisme du sol

<400> 92  
 gtctgtgtagc aataccagat gatggcaagt ggcggaacgg tgagtaatac gtagggatct 60  
 gccacagaaga gggggacaac ccggggaaac tcggggctaact accgcatact attctgagga 120  
 aaaaagcttg gcgcaagcca ggcgcttttg gaggaaccta cgtccgatta gctagtgtgt 180  
 gaggtaaagg ctacccaagg cagagatcgg tagctgggtc gagaggatga ccagccacac 240  
 tgggactgag acacggccca gactcctacg ggaggcagca gtggggaata ttggacaatg 300  
 ggggcaaccc tgatccagcg atgcgcgctg tgtgaagaag gccttcgggt tgtaaagcac 360  
 tttagtgtgg gaagaagtaa tgttttttaa tagagagcat tgttgacggt acccaagaa 420  
 taagcaccgg ctaactctgt gccagcagcc gcggtaatac agaggggtga agcgttaac 480  
 ggagtacttg ggcgtaaaagg gcgcgtaggc ggtgttgcaa gtgagatgtg aaatccctgg 540  
 gcttaacccta ggaaccgcat tttagactgc aatgctagag tacagttagg ggtagtggaa 600  
 ttccgggtg agcggtgaaa tgcgtagaga tcggaaggaa caccagtggc gaaggcgact 660  
 acctggactg acactgacgc tgaggcgaga gagcgtgggg agcaaacagg attagatacc 720  
 ctggtagtcc acgctgtaaa cgtatgagaac tagatgttgg tgcgcgcgag cgcacaagta 780  
 tcgaagctaa cgcgataagt tctccgcctg gggagtagc cgcgaaggtt aaaactcaaa 840  
 ggaattgacg ggggcccgcga caagcggtgg agcatgttgt ttaattcgat gcaacgcgaa 900  
 gaaccttacc tacccttgac atccacagaa tttgatagag atatcgaagt gccgaaggaa 960  
 actgtgagac aggtgctgca tggctgtcgt cagctcgtgt tgtgagatgt tgggttaagt 1020  
 ccgtaaacgg gcgcgaacct tatccttagt tgccaacacg taatggtggg gactctaagg 1080  
 agactgcggc tgaagaaccg gaggaagggt gggacgacgt caagtcatac tggcctttat 1140  
 gggtagggct acacacgtgc tacaatgggg cgtacagagg gttgccaaac tgcgaagggg 1200  
 agccaatccc ggaaagcgcc tcgtagtcca gattgaagtc tgcaactcga 1250

<210> 93  
 <211> 1545  
 <212> ADN  
 <213> Organisme Inconnu  
 <220>  
 <223> Origine de la séquence :Organisme du sol

<400> 93  
 ccaggaaaaca gctatgacca tgattacgcc aagcttggtg ccgagctcgg atccactagt 60  
 aacggccgcc agtgtgtctg aattcgccct tcaggcctaa cacatgcaag tcgaacggca 120  
 gcacagggga gcttgcctcc tgggtggcga gtggcggacg ggtgaggaat acatcggaat 180  
 ctgcccagtc gtgggggata acctcgggaa accgggacta ataccgcata cgaccttagg 240  
 gtgaaagcgg aggaaccgcaa ggcttcgcgc gattggatga gccgatgtcg gattagcttg 300  
 ttggcggggg aacggccccc caaggcgacg atccgtagct ggtctgagag gatgatcagc 360  
 cacactggaa ctgagacacg gtccagactc ctacgggagg cagcagtgagg gaattattga 420  
 caatggcgcc aagcctgatc cagccatgcc gcgtgagtga agaaggcctt cgggttgtaa 480  
 agctcttttg tccggaaaaga aaagcttttc gttaataccc ggaagctcctg acggtaccgg 540

```

aagaataagc accggctaac ttctgtgccag cagccgcggg aatacgaagg gtgcaagcgt 600
tactcgggaat tactggggcgt aaagcgtgcg taggtgggtt gttaaagtcgt atgtgaaagc 660
cctgggcgtca acctgggaat tgcaactggat actggcaggc tagagtgcgg tagaggatgg 720
cggaattccc ggtgtagcag tgaatgcgt agagatcggg aggaacatct gtggcggaagg 780
cgcccatctg gaccagcact gacactgagg cacgaaagcg tggggagcaa acaggattag 840
atacctcgtt agtccagcgc ctaaacgatg cgaactggat gttgggagca actaggctct 900
cagtatcgaa gctaaccgct taagtctgcc gcctggggag tacggtcgca agactgaaac 960
tcaaagggaat tgacgggggc ccgcacaagc ggtggagtat gtggtttaat tcgatgcaac 1020
gcgaagaacc ttacctggcc ttgacatcca cggaacttac cacagatggt ttggtgcctt 1080
cggnaaccgt gagacagggt ctgcatggct gtctgcagct cgtgtcgtga gatgttgggt 1140
taagtccgcg aacgagcgca acccttgctc ttagttgcca gcacgtaatg gtgggaactc 1200
taaggagact gccggtgaca aaccggagga agtggggatg gacgtcaagt catcatggcc 1260
cttacggcca gggctacaca cgtactacaa tggtcgggtac agagggttgc aaagccgcga 1320
ggtagagcca atcccagaaa accgatccca gtccggatcg aagtcctgaa ctgcacttcg 1380
tgaagtcgga atcgctagta atcgcggatc agaatgccgc ggtgaatacg ttcccgggcc 1440
tgttaccata cgcccgaagg cgaattctgc agatattcat cacactggcg gccgctcgag 1500
catgcatcta gagggcccaa ttccgccctat agtgagtcgt attac 1545

```

&lt;210&gt; 94

&lt;211&gt; 1549

&lt;212&gt; ADN

&lt;213&gt; Organime Inconnu

&lt;220&gt;

&lt;223&gt; Origine de la séquence :Organisme du sol

&lt;400&gt; 94

```

ttttaaacg acggccagtg aattgtaata cgactcacta tagggcgaat tggggccctct 60
agatgcatac tcgagcggcc gccagtgtag tggatatctg cagaattcgc ccttcaggcc 120
taacacatgc aagtcgagcg gcagcgcggg gcaacctggc ggcgagcggc ggacgggtga 180
ggaatgcctc ggaatctacc ctgtcgtggg ggataacgta ggcgaactta cgctaatacc 240
gcatacgacc gagagtgtaa agtgggggac cgcaaggcct cagcgcgatg gatgagccga 300
tgccggatta gctagtgtgt gaggtaaaag ctacccaagg cgacgatccg tagctggctc 360
gagaggatga tcagccacat tgggactgag acacggccca aactcctacg ggaggcgaga 420
gtggggaata ttggacaatg ggccgaagcc tgatccagcc atgccgcgtg tgtgaagaag 480
gccttcgggt tgtaaaagac tttgttcgg gaagaaatcg tgcgggttaa taccagatc 540
ggatgacggg accgaagaag taagcacagg ctaactctgt gccagcagcc ggggtaatac 600
gaagggtgca agcgttactc ggaatcactg ggcgtaaaag gtgcgtaggc ggttggttaa 660
gtctgctgtg aaagccctgg gctcaacctg ggaactgcag tggatactgc ccagctagag 720
tgtgatagag gatggtggaa ttcccgggtg agcgggtgaa tgcttagaga tcgggaggaa 780
caccagtggc gaaggcggcc atctggatca acactgagcg tagaggcaga aagcgtgggg 840
agcaaacagg attagatacc ctggtagtcc acgccctaaa cgaatgcgaa tggacgttgg 900
gagcaacttg gctctcagtg tcgaagctaa cgcgctaagt tcgcgcgctg gggagtagcg 960
tcgcaagact gaaactcaaa ggaattgacg tgggcccgcga caagcgggtg agtatgtgtg 1020
ttaattcgat gcaacgcgaa gaaccttacc tggcctgtac atccacggaa cttaaccagag 1080
atggtttggt gccttcggaa ccgtgagaca ggtgctgcat ggcgtgtcgt agctcgtgtc 1140
gtgagatgct ggggttaagt ccgcaacgag cgcaaccctt gtcccttagtt gccagcagc 1200
aatggtggga actctaagga gactgccggt gacaaacccg aggaaggtgg ggatgacgtc 1260
aagtcatcat ggcctctacg gccagggtca cacacgtact acaatggctg gtacaagagg 1320

```



```

gttgcaaacg ccgcgaggta gagccaatcc cagaaaaccc gatcccagtc ccggatcgaa 1380
gtctgcaact cgacttcgtg aagtcggaat cgctagtaat cgcggatcag aatgccgcgg 1440
tgaatacgtt cccgggcctt gtacacaccg cccaagggcg aattccagca cactggcggc 1500
cgttactagt ggatccgagc tcggtacca gcttggcgta atcatggtc 1549

```

&lt;210&gt; 95

&lt;211&gt; 1276

&lt;212&gt; ADN

&lt;213&gt; Organisme Inconnu

&lt;220&gt;

&lt;223&gt; Origine de la séquence :Organisme du sol

&lt;400&gt; 95

```

ctggcggcga gcggcggagc ggtgaggaa acatcggaat ctaccagtc gtgggggata 60
acgtaggaaa acttacgcta ataccgcata cgacctgagg gtgaaagcag gggatcgcaa 120
gaccttgccg gattggatga gccgatgtcc gattagctag ttggtgagg aaaggtcac 180
caaggcgacg atcggtagct ggtctgagag ggtgatcagc cacactggaa ctgagacacg 240
gtccagactc ctacgggagg cagcagtggtg gaatttgga caatgggcgc aagcctgatc 300
cagcctatgcc gcgtgtgtga agaaggcctt cgggtgtgaa agcacttttg ttcggaaga 360
aatcttcoga gttaatacct cgggaggatg acggtaccgg aagaataagc accgctaacc 420
ttcgtgccag cagccgcggt aatacgaagg gtgcaagcgt tactcggaat tactgggcgt 480
aaagcgtgag taggtggttc gttaagtctg ccgtgaaagc cccgggctca acctgggaat 540
tgcggtggat actggcgagc tagagtgcgg tagagggtgg tggaaattccc ggtgtagcag 600
tgaaatcgct agagatcggg aggaacatct gtggcgaagc ggccacctgg accagcactg 660
acactgaggg acgaaagcgt ggggagcaaa caggattaga taccctggta gtccacgcc 720
taaacgatgc gaactggagc ttgggagcaa ctaggctctc agtgtcgaag ctaacgcgtt 780
aagttcgccg cctggggagt acggtcgcaa gactgaaact caaaggaatt gacgggggcc 840
cgcacaacgg gtggagtgtg tggtttaatt cgatgcaacg cgaagaacct tacctggcct 900
tgacatccac ggaatccttt agagatagag gagtgccttc gggaaccgtg agacaggtgc 960
tgcatggctg tcgtcagctc gtgtcgtgag atgttgggtt aagtcgccga acgagcgcaa 1020
cccttgtcct tagttgccag cgcgtaatgg cgggaacctct aaggagactg ccggtgacaa 1080
accggaggaa ggtggggatg acgtcaagtc atcatgccc ttacggccag ggctacacac 1140
gtactacaat ggtggggaca gagggtcgcg aagccgcgag gtggagccaa tccagaaaac 1200
cccactcag tcgggatcgg agtcgtcaac tcgactccgt gaagtcgga tcgctagtaa 1260
tcgcggtcag catgcc 1276

```

&lt;210&gt; 96

&lt;211&gt; 1306

&lt;212&gt; ADN

&lt;213&gt; Organisme Inconnu

&lt;220&gt;

&lt;223&gt; Origine de la séquence :Organisme du sol

&lt;400&gt; 96

```

cagggatcag tagagtggca aacgggtgag taacgcgtgg gcgacctacc ttcgagtggg 60
ggataacctt ccgaaggagg ggctaatacc gcatgacatc ccgtgtttgg atacacggac 120

```

```

atcaaagccg gggatcgcaa gacctggcgc ttggagaggg gcccgcgctcc gattagctag 180
ttggtgaggt caccggctcac caaggctccg atcggtatcc ggcttgagag ggcggacgga 240
cacactggga ctgagacacg gcccgagactc ctacgggagg cagcagctggg gaattgtctg 300
caatggggcg aagcctgacg acgcaacgcc gcgtggagga tgaagacctt cgggtcgtaa 360
actcctttcg accgagatga agacccggcg gcctaatacg ccggcggtatt gacagtatcg 420
agggagaag ccccgctcaa ctccgtgccg gcagccggcg taatacgggg ggggcaagcg 480
ttgttcggaa ttactggggc taaagggttc gtagggtgct cgctaagtca gacgtgaaa 540
ccctcaagctc aactggggaa ctgcgtctga gactggcaag cttgagtgca ggagaggaac 600
gcggaattcc aggtgtagcg gtgaaatcg tagatatctg gaggaacacc ggtggcggaag 660
gcggcgctct ggaactgcaac tgacactgag gaacgaaagc taggggagca aacgggatta 720
gataccccgg tagtctagc cctaaacgat gaatgcttgg ttggcggggt atcgatccct 780
gcggtgcccg agttaacgag ataagcattc gcctggggg gtacgggtcg aaggctgaaa 840
ctcaaaaggaa ttgacggggg cccgcacaag cgggtggagca tctggttcaa ttcgacgcaa 900
cggaagaac cttacctagg ctccaagtgc agatgacctc cggtgaaagc cgactttcgc 960
aagaacatct gttagggctg tgcattggct tcgtcagctc gtgtcgtgag atgttggggt 1020
aagtcccgca acgagcgcaa cccttgtttc ctgttgccat cagggttaagc tgggcactct 1080
ggagagactg ccggtgacaa accggaggaa ggtgggggat acgtcaagtc agcatggcct 1140
ttatgtctag ggctacacac gtgctacaat ggccggtaca aagccgtcga accccggag 1200
ggtgagccaa tcgcagaaag ccggtctcag ttccgatagc aggctgcaac tcgcctgctt 1260
gaagttggaa tcgctagtaa tcgcggatca gcattgccgc gtgaat 1306

```

&lt;210&gt; 97

&lt;211&gt; 1300

&lt;212&gt; ADN

&lt;213&gt; Organisme Inconnu

&lt;220&gt;

&lt;223&gt; Origine de la séquence :Organisme du sol

&lt;400&gt; 97

```

cccgacgggt gagtagatgg caaacgggtg agtaaacact gggtagacct cctcagagtg 60
ggggataaac acccgaaaag gtcgctaata ccgcataaca tctgtctttt ggatagacgg 120
agatcaaaag cggggatcgc aagacctggc gcttagagag gggcccgccg ggcttagct 180
agttggtgag gtaacggctc accaaggcaa cgatcggtat ccggccttag agggcgagcg 240
gacactactg gactgagaca cgcccagac tcctacggga ggcagcagtg gggaaattgt 300
cgcaatgggc gcaagcctga cgaacgcaac ccgcgtggag gatgaagatc ttccggtcgt 360
aaactccttt cgatcgggaa gaacgcctct ggtgtgaaca coactcagagg gtgacggtac 420
cgagagaaga agccccggct aactctgtgc cagcagccgc ggtaatacag gggggggcaag 480
cgtgtgtcgg aattactggg cgtaaaaggcg tcgtaggcgg ccgcttaagt ccgactgtaa 540
atccccagcg ttaactcggg aactgcgtcg gatactggcg ggcttgaatc cgggagaggg 600
atgcggaatt ccagggtgag cggtgaaatg ctgagatatc tggaggaaca ccggtggcga 660
aggcggcatc ctggaccggg attgacgctg aatagcgaaa gccagggggag caaacgggat 720
tagatacccc ggtagtctcg gccctaaacg atgaatgttt ggtgtggcgg gtatcgatcc 780
ctgcgctgcc gaagctaacg cattaaacat tccgcctggg gagtacggtc gcaaggctga 840
aactcaaaag aattgacggg gggccgcaca agcgggtggag catgtggttc aattcgacgc 900
aacgcgaaga accttaacca ggctcgaacg gcattggaca tccggcgaaa gccggtctcc 960
gcaaggggcg atgtcgaggg gctgcattgc tgcgtcagc tcgtgtcttg agatgttggg 1020
ttaagtcccc caacgagcgc aacccttgtc cgctgtgccc atcacgttat ggtgggact 1080
ctcggagac tgcgggtgat aaaccggagg aaggtgggga tgacgtcaag tcacgatggc 1140

```

```

ctttatgtctt ggggctacac acgtgctaca atggccggta caaaccggtg cgatctcgca 1200
agagtgaagct aatcggagaa agccgggtctc agttcggatt gcaggctgca actcgcctcg 1260
atgaagttagg aatcgctagt aatcggcgat cagcacgccc 1300

```

```

<210> 98
<211> 1233
<212> ADN
<213> Organisme Inconnu

```

```

<220>
<223> Origine de la séquence :Organisme du sol

```

```

<400> 98
acggagcgcc agacggggaga gtaacacgtg ggaacgtgcc ctttggttcg gaacaacaca 60
gggaaacttg tgctaatacc ggataagccc ttacggggaa agatttatcg ccaaaggatc 120
ggcccgcgctc tgatttagcta gttggtgagg taacggctca ccaaggcgac gatcagtagc 180
tggtctgaga ggaatgatcag cctcaactggg actgagacac ggcccagact cctacgggag 240
gcagcagttg ggaatatattg acaatggggc caagcctgat ccagccatgc cgcgtggatg 300
atgaaggccc tagggttgta aagtcctttc ggccggggaag ataatgacgg tacccgcaga 360
agaaagcccc gctaacttcg tgccagcagc cgcggttaata cgaagggggc tagcgttgct 420
cggaataact gggcngtaaa gcgcacgtag gcggcttttt aagtcagggg tgaatcctg 480
gagctcaact ccagaactgc ctttgatact gagaagcctg agtcggggag aggtgagtgg 540
aactgcgagt gtagaggtga aattcgtaga tattcgcaag aacaccagtg gcgaaggcgg 600
ctcactggcc cggtactgac gctgaggtgc gaaagcgtgg ggagcaaaac ggattagata 660
ccctggtagt ccacgctgta aacgatggat gctagcgtt gtccgggtta ctcgtagtg 720
gcgcagctaa gcgattaagc atcccgcctg gggagtagcgc tcgcaagatt aaaactcaaa 780
ggaattgacg ggggcccgcga caagcgggtg agcatgtgt tcaattcgaa gcaacgcgca 840
gaaccttacc agcccttgac atgtcccga tgagtaccag agatggaact cttcagttcg 900
gctggcggga acacaggtgc tgcattggctc tcgtcagctc gtgtcgtagc atgttggggt 960
aagtcgccga acgagcgcaa ccctcgccct tagttgccat catttagttg ggcactctaa 1020
ggggactgcc ggtgataagc cgcgaggaa gctgggatga cgtcaagtc tcatggccctc 1080
tacgggctgg gctacacacg tgctacaatg gcggtgacag tgggatgcag aggggtaacc 1140
ccgagcaaat ctcaaaaagc cgtctcagtt cggattgtgc tctgcaactc gagcacatga 1200
agttggaatc gctagtaatc gcagatcagc acg 1233

```

```

<210> 99
<211> 1304
<212> ADN
<213> Organisme Inconnu

```

```

<220>
<223> Origine de la séquence :Organisme du sol

```

```

<400> 99
cgaaatcccc cagggatcag tagagtggca aacgggtgag taacacgtg gtgacctgcc 60
ttcgactggg ggataacgtc ccgaaaggga cgctaatacc gcataacatc ctgctcttga 120
acgagtggag atcaaaagctg gggatcgcaa gacctagcgc tcaaaaggag gccccgcctc 180
gattagctag ttggtggggt aacggctcac caaggcgagc atcagtatcc ggcctgagag 240

```

```

ggcgagacgga cactctggga ctgagacacg gccagacgtc ctacggggagg cagcagtggtg 300
gaattgtctcg caatggggcgc aagcctgacg acgcaacgcc gcgtggaggga tgaagatctt 360
cgggtcgtaa actcctcttcg atcgagacga acggctctcg ggtgaacaaat ccggaggagt 420
gacggtaccg agagaagaag ccccggtctaa ctccgtgcca gcagccgcgg taatacgggg 480
ggggcaagcg ttgttcggaa ttactgggctg taaagggtctc gtaggcgccc aactaagtca 540
gacgtgaat cctctcggtct aacccgggaa ctgcgtctga tactggatgg ctagaggttg 600
ggagagggat gcggaattcc aggtgtagcg gtgaaatgcg tagatatctg gaggaacacc 660
ggtggcgaag gcggcatcct ggaccaattc tgacgctgag gagcgaaagc caggggagca 720
aacgggatta gatacccccg tagtctctggc cctaaacgat gaatgcttgg tctggcggtg 780
atcgatccct gccgtgccga agctaacgca ttaagcatct cgctggggga gtacggtcgc 840
aaggctgaaa ctcaaaggaa ttgacggggc ccgcgacaag cggtggagca tgtggttcaa 900
ttcgacgcaa cgcgaagaac cttaccagg cttgaacagc gagtgaaccac tcttgaaaag 960
gagcttccgc aaggacactc gtagaggtgc tgcattggctg tctgcagctc gtgtcgtgag 1020
atgttgggtt aagtcgccga acgagcgcaa ccttgttctg ctgttgccat cagcttatgg 1080
tgggcactct gcaaagactg ccggtgataa accggaggaa ggtggggatg agctcaagtc 1140
agcatggcct ttatgtctcg ggctacacac gtgtacaat ggccggatga aaccgtcgca 1200
aaaccgttaag gtcgagctaa tcggagaaag ccggtctcag ttcggatcgt cgggtgcaac 1260
tcgccggcgt gaagttggaa tcgctagtaa tcgcgatca gcac 1304

```

&lt;210&gt; 100

&lt;211&gt; 1197

&lt;212&gt; ADN

&lt;213&gt; Organisme Inconnu

&lt;220&gt;

&lt;223&gt; Origine de la séquence :Organisme du sol

&lt;400&gt; 100

```

tctagtgcgc caccgggtgcg taacgcgtgg gaatctgccc ttgggttcgg gataaacagt 60
ggaaaacgct gctaatacgc gatgatgtct tcggaccaaa gatttctcgc ccagggatga 120
gcccgctcgc gattagctag ttggtgaggt aaaggctcac caaggcgacg atccgtagct 180
ggtctgagag gatgactcag cactctggga ctgagacacg gccagactc ctacgggagg 240
cagcagtggtg gaattattgga caatggggcga aagcctgacg caccatggcc gcgtgagtga 300
tgaaggcctt agggttgtta agctcttttg ccgggatga taatgacagt accgggagaa 360
taagccccgc ctaactccgt gccagcagcc gcggtaatca ggagggggct agcttgttct 420
ggaattactg gccgtaaagc gcacgtaggg ggctttgtta gttagaggtg aaagcccgga 480
gctcaactcc ggaactgcct ttaagactgc atcgcttgaa cgtcggagag gtaagtggaa 540
ttccgagtgt agaggtgaaa ttcgtagata ttccgaaaga caccagtgcc gaaggcgact 600
tactggacga ctgttgacgc tgaggtgcga aagcgtgggg agcaaacagg attagatacc 660
ctggtagtcc acccggtaaa cgtatgatgc tagctgtcgg ggctcatgga gtttcggtgg 720
cgcagctaac gcgttaagtc atccgcctgg ggaagtacggc cgcaaggtta aaactcaagg 780
aaattgacgg gggcctgcac aagcgggtgga gcatgtggtt taattcgaag caacgcgcag 840
aaccttacca gcgtttgaca tggtaggacg gtttccagag atggattcct tcccttacgg 900
gacctacaca caggtgctgc atggctgtcg tcagctcgtg tctgtgagatg ttgggttaag 960
tcccgcaacg agcgcaaccc tcgtcttttag ttgctaccat ttagtgtggc actctaaaga 1020
aactccgggt gataagccgg aggaagggtgg ggaatgacgtc aagtcctcat ggccttacg 1080
cgctgggcta cacacgtgct acaatggcgg tgacagtggt cagcaaaact gcgagagtga 1140
gcaaatcccc aaaaaccgtc tcagtttcgga ttgttctctg caactcgaga gcatgaa 1197

```

<210> 101  
 <211> 1352  
 <212> ADN  
 <213> Organisme Inconnu

<220>  
 <223> Origine de la séquence :Organisme du sol

<400> 101  
 cgacggccag tgaattgtaa tacgactcac tatagggcga attggggccct ctatagtcac 60  
 gctcgagcgg ccgccagtgat gatggatata tcgagaattc gcccttcagg cctaaccacat 120  
 gcaagtgcga cgagaaaggg cttcgggccc ggtagcagtg cgacggggtg agtaaacagt 180  
 aggcaatctc cctcagtgat gtggataaac ttccgaaagg agggctaata cagcatgaga 240  
 ccacgagctc gcagagcttg tggccaaagc ggacctcttc ttgaaagttc gcgcttgagg 300  
 atgagcctcg ggcccatcag ctagtggta gggtaatggc ctaccaaggc taagacgggt 360  
 agctggtctg agaggacgga cagccacact ggaactgaga cagcggtccag actcctaagg 420  
 gaggcagcag tggggaaatct tgcgcaatgg acgaaagtct gacgcagcga cgcccggtga 480  
 gogatgaagg ccttcggggt gtaaagctct gtggggagag acgaataagg tgcagctaat 540  
 acctgcatcg atgacgggat ctcttagca agcaccggct aactctgtgc cagcagccgc 600  
 ggtaagacag aggggtgcaa cgttgttcgg aattactggg cgtaaaagcg gtgtaggcgg 660  
 ctgtgtaatg cgggcgtgaa atcccatggc tcaacatgg aagtgcaccc gaaactgcgt 720  
 agctagatgc ctggagagga aggtggaatg cttggtgtag aggtgaaatt cgtagatata 780  
 aagcggaaca ccggtggcga agcggccttc tggacagtga ctgacgtga gacgcgaaag 840  
 cgtggggagc aaacaggatt agataccctg gtatgccag ccgtaaacga tgaatgctag 900  
 agctgggggt gcatgcactt cgggtgcgcc gctaacgcgt taagcattcc gctggggag 960  
 tacggccgca aggttaaaac tcaaaggaat tgacgggggc ccgcacaagc ggtggagcat 1020  
 gtggtttaat tcgaagcaac gcgcaaacct taccacacct tgacatgtcc attgccggtc 1080  
 cgagagattg gaccttcagt tcggctggat ggaacacagg tgctgcatgg ctgtcgtcag 1140  
 ctctgtctgt gagaatgttg gttaagtccc gcaacgagcg caaccctac cgccagtttg 1200  
 catcattcag ttgggcactc ttggtggaact gccggtgaca agccggagga agcggggatg 1260  
 acgtcaagtc ctatgggcc ttatgggttg ggctacacac gtgctacaat ggcggtgaca 1320  
 gtgggacgcg aagtccaaga tggacaaatc cc 1352

<210> 102  
 <211> 1361  
 <212> ADN  
 <213> Organisme Inconnu

<220>  
 <223> Origine de la séquence :Organisme du sol

<400> 102  
 aacagctatg accatgatta gcccaagctt ggtacgagc tcggatccac tagtaacggc 60  
 cgccagtggt ctggaattcg cccttcaggc ctaacacatg caagtgcga ggaatcctcg 120  
 ggattagtgg cggaagggtg agtaaacagt ggaacgtgc cctttgggtc ggaacaactc 180  
 agggaaactt gacataatc cggataagcc ttccgagga aagatttat gccattggag 240  
 cggcccgctg aggatagct agttggtgag gtaaaagctc accaaggcga cgaatcctag 300  
 ctggtctgag aggatgatca gccacattgg gactgagaca cgcccaaac tcctacggga 360

```

ggcgagcagtg ggggaatcttg cgcaatgggc gcaagcctga tccagccatg ccgcgtgagt 420
gatgaaggcc ttaggggttg aaagctcttt caccggagaa gataatgacg gtatccggag 480
aagaagcccc ggctaacttc gtgccagcag ccgcggtaat acgaaggggg ctgacgcttg 540
tcggaattac tgggcgtaaa gcgcacgtag gcggtatatt aagtcagggg tgaaaatcca 600
gagctcaact ctggaactgc ctttgatact ggggtatctg agtatggaa aggttaagtgg 660
aattccgagt gttagaggtg aattcgtaga tattcggagg aacaccagtg gcgaaggcgg 720
cttactggtc cttactgac gctgaggtgc gaaagcgtgg ggagcaaca ggattagata 780
ccctggtagt ccacgcctga aacgatgaat gttagccgtc gggcagtata ctgttcggtg 840
gcgcagctaa cgcattaaac attccgcctg gggagtagcg tcgcaagatt aaaactcaaa 900
ggaattgacg ggggcccgcg caagcgggtg agcatgtggt ttaattcgaa gcaacgcgca 960
gaaccttacc agctcttgac attcgggggt tgggcagtg agacattgtc cttcagttag 1020
ctgggcccc gaacagggtg tgcattggctg tcgtcagctc gtgtcgtgag atgttgggtt 1080
aagtcgccga acgagcgcaa ccctcgccct tagttgccag catttagttg ggcactctaa 1140
ggggactgcc ggtgataaag cgagaggaag gtggggatga cgtcaagtc tcattggcct 1200
tacgggctgc gctacacacg tgctacaatg gtgggtacag tgggcagcga gacagcgatg 1260
tcgagctaag ctccaaagcg catctcagtt cggattgcat ctgcaactcg agtgcagtga 1320
gttggaatcg ctagtaatcg cagatcagca tgcgtcggtg a 1361

```

&lt;210&gt; 103

&lt;211&gt; 1300

&lt;212&gt; ADN

&lt;213&gt; Organisme Inconnu

&lt;220&gt;

&lt;223&gt; Origine de la séquence :Organisme du sol

&lt;400&gt; 103

```

catgttttagt agcaataacta aatgatgacg agcggcggac ggggtaggaa cacgtaggaa 60
cctgcaccaag agaggggggac aaccaaggga aactttggct aataccgcgt aatctctacg 120
gagaaaagtt gcccgtaag gtggcgcttt tggaggggcc tcgctccgat tagttagttg 180
gtgaggtaat agctaccac gactgtgatc ggtaactggt ctgagaggac gaccagtcac 240
actgggactg agacacggcc cagactccta cgggaggcag cagtggggaa tcttggacaa 300
tgggggcaac cctgatccag cgtatgcccg tgggtgaaga aggccttcgg gttgtaaagc 360
ccttttaggcg ggggaagaag atatgggatg aataagcctg tattttgacg gtaccgcgag 420
aataagcacc ggcaaaactc gtgccagcag ccgcggtaat acagaggggt cgagcgtaa 480
tcggatttac tgggcgtaaa gggcgcgtag gcggttgtgt gagtgtgatg tgaagggccc 540
gggctcaacc tgggaagtgc atcgcaaacg acacaactgg agtatatgat aggggtggcg 600
aatttccggt gttagcggtg aatgcgtaga gatcggaagg aacgtcgatg gcgaaggcag 660
ccacctggca taatactgac gctgaggcgc gaaagcgtgg ggagcgaaca ggattagata 720
ccctggtagt ccacgcctga aacgatgaga actagatggt ggagggggaa cccttcagta 780
tcgaagctaa cgcgataagt tctccgcctg ggaagtacag tcgcaagact gaaactcaaa 840
agaattgacg gggggccgcg caagcgggtg agcatgtggt ttaattcgat gcaacgcgaa 900
gaaccttacc taaccttgac atctcgcaaa tcttggcgag aggtgagagt gccgaagga 960
gcgcagagac aggtgctgca tggctgtcgt cagctcgtgt tgtgagatgt tgggttaagt 1020
cccgtaacga gcgcaacctt tgtctctagt tgccatcatt tagttgggga ctctaaaggag 1080
accgcggtg atgaaccgga ggaaggcggg gacacgctca agtcatcatg gcctttatgg 1140
gtagggctac acacgtgcta caatggggcg tacagagggt cgccaacccg cgagggggag 1200
ccaatctctt aaagcgtctc gtagtccgga ttggagtctg caactcgact ccatgaagtc 1260
ggaatcgcta gtaatcgcg atcagcagtg ccgcggtgaa 1300

```

<210> 104  
 <211> 1250  
 <212> ADN  
 <213> Organisme Inconnu

<220>  
 <223> Origine de la séquence :Organisme du sol

<400> 104  
 tgtagcaata catcagtggt agacgggtga gtaacacgtg ggaaccttcc tcgttgtagc 60  
 ggacaactca gggaaacttg agctaatacc gtatacgtcc gagaggagaa agatttatcg 120  
 caatgagacg ggcgcgcgtc ggattagcta gttggttaag taacggctta ccaaggcgac 180  
 gatccgtagc tgatctgaga ggatgatcag ccacactggg actgagacac ggcccagact 240  
 cctacggggag gcagcagtggt ggaatcttgg acaatgggag caagcctgat ccagccatgc 300  
 cgcgtagtag aagaaggcct taggggttga aagctctttt gccaggggagc ataatgacgc 360  
 tacctgagaa taagccccgg caaacttcgt gccagcagcc gcggttaatac gaagggggct 420  
 agcgttgttc ggatttactg ggcgtaaagc gcacgtaggc gggctcgttaa gtcagggggtg 480  
 aaatcccgga gctcaactcc ggaactgcct ttgatactgg cgaccttgag gctggaagag 540  
 gttagtggaa ttcccagtg agaggtgaaa ttctgtagata ttgggaagaa caccagtggc 600  
 gaagggcggt aactgggtcca gatctgacgc tgaggtgcga aagcgtgggg agcaaacagg 660  
 attagatacc ctggtagtcc acgcgtgtaa ctatgggtgc tagctgtcag cgggcttgct 720  
 cggtgggtgc gcagctaacg cattaagcac ccgcctcggg gagtacggtc gcaagattaa 780  
 aacttaagg aattgacggg ggcccgacac agcgggtggag catgtggttt aattcgaagc 840  
 aacgcgcaga accttaccac cccttgacat ccgcatcgcc gacaccagag atggagtctc 900  
 tcagttcgcc tggatcgagg acaggtgctg catggctgtc gtcagctcgt gtcgtgagat 960  
 gttgggttaa gtcccgcaac gagcgcaacc ctgcctttaa gttgccatca tttagtgggg 1020  
 cactctaagg ggactgcggg tgataagccg gaggaagggt gggatgacgt caagtctcca 1080  
 tggcccttac ggggtgggct acacacgtgc tacaatggcg gtgacaatgg gcagctactt 1140  
 cgcaaggaga agctaatacc aaaaagccgt ctcagttcag attgcaactg gcaactcggg 1200  
 tgcataagat tggaaatcgt agtaatcgt aatcagcagg tagcgggtgaa 1250

<210> 105  
 <211> 1302  
 <212> ADN  
 <213> Organisme Inconnu

<220>  
 <223> Origine de la séquence :Organisme du sol

<400> 105  
 ggcttcggct ccccggtaga gtggcgagc ggtgagtaac acgtgggttaa tctgcctttg 60  
 ggtggggaaat aaccctctga aagaggggct aataccgcat aacgcagcgg caccgaatgg 120  
 tgacagtgtg taaagtgggg gatcgcaaga cctcagcgct gaagaggagc ccgcgccga 180  
 tttagctagt ggtgcggttaa tggcgtaacca aggcggcgat cggtagccgg cctgagaggg 240  
 cggacggcca cactggcact gagagacggg ccagactcct acgggaggca gcagtgggga 300  
 attttgggca atgggcgcaa gcctgaccca gcaacgcgc gtgaaggagc aaatccctct 360  
 gggatgtaaa cttcgaagat tggggaagaa atccgtgtga ggataatgca caccgggatga 420

```

cggtacecaa cgtaagcccc ggctaactac gtgccagcag ccgcggtaat acgtaggggg 480
caagcgttgt tcggaattac tgggcgttaa gggcgcgtag cgcgtacgac aagctcggag 540
tgaagccccc gggctcaacc ccggaatgtc tttggaaact gtcgaaactg agtcggaag 600
aggcatctgg aattccagct gtacgcgtga aatcgtaga tattggggaag aacacctgag 660
gcgaagcgcg gatgctgggc cgacactgac gctgaggcgc gaaagccagg ggagcgaaag 720
ggattagata ccccggtagt cctggcccta aacgatggat acttggtgtg tggggttctc 780
gaagtccccc cgtgccggag ctaacgcggt aagtatcccg cctggggagt acggtcgcaa 840
ggctgaaact caaaggaatt gacggggacc cgcacaagcg gtggagcatg tggttcaatt 900
cgacgcaacg cgaagaacct tacctgggtt aaatcctacc tcgtcgccct agagatgagg 960
tttcccttcg ggggaggtag gacggtgctg catggtctgc tcagctcgt gccgtgaggt 1020
gttgggttaa gtcccgcac gacgcgaacc cttaccacta gttgccagcg gttcgccgg 1080
gcactctatt gggactgccg gtgacaaacc ggaggaaggt ggggatgacg tcaagtcatc 1140
atggccttta tgtccagggc tacacacgtg ctacaatggc cggaacaaag cgcagcaaac 1200
ccgcgagggg gagccaatcg caaaaatccg gtctcagttc gggattggagt ctgcaactcg 1260
actccatgaa gttggaatcg ctagtaatcg cggatcagca tg 1302

```

&lt;210&gt; 106

&lt;211&gt; 1281

&lt;212&gt; ADN

&lt;213&gt; Organime Inconnu

&lt;220&gt;

&lt;223&gt; Origine de la séquence :Organisme du sol

&lt;400&gt; 106

```

tgcttctctt gagagcggcg gacgggtgag taatgcctag gaactcgcct ggtagtgggg 60
gataacgttc ggaacaggac gctaataacc catacgtcct acgggagaaa gcaggggacc 120
ttcgggcctt gcgctatcag atgagcctag gtccgattag ctagtgtggt aggtaatggc 180
tcaccaaggc gacgatccgt aactggtctg agaggatgat cagtcacact ggaactgaga 240
cacggtccac actcctacg gaggcagcag tggggaatat tggacaatgg gcgaaagcct 300
gatccagcca tgcgcggtgt gtgaagaagg tcttcggatt gtaaagcact ttaagtgtga 360
aggaagggtc gtaaattaat actttgtctg tttgacgtta ccgacgaat aagcaccggc 420
taactctgtg ccagcagccg cggtaataca gagggtgcaa gcgttaatcg gaattactgg 480
gcgtaaagcg cgcgtaggtg gtttgttaag ttggatgtga aatccccggg ctcaacctgg 540
gaactcgatt caaaactgac tgactagagt atggtagagg gtggtggaat ttctgtgta 600
gcggtgaaat cgttagatat aggaagggaac accagtggcg aaggcgacca cctggactaa 660
tactgacact gaggtgcgaa agcgtgggga gcaaacagga ttagataccc tggtagtcca 720
gcgcgtaaac gatgtcaact agcgttgga agccttgagc ttttagtggc gcagctaacc 780
cattaagtgt accgcctggg ggttacggcc gcaagggttaa aactcaaatg aattgagggg 840
ggcccgacac agcgggtggag catgtggttt aattcgaagc aacgcgaaga acctaccag 900
gcttgacat ccaatgaact ttctagagat agattggtgc cttcgggaac attgagacag 960
gtgctgcatg gctgtcgtca gctcgtgtcg tgagatgttg ggttaagtcc cgtaacgagc 1020
gcaacctctg tccttagtta ccagcagcag atggtgggca ctctaaggag actgccggtg 1080
acaaaccgga ggaaggtggg gatgacgtca agtcatcatg gccttacgg cctgggtcac 1140
acacgtgcta caatggtcgg tacagagggg tgccaagccg cgaggtggag ctaatccac 1200
aaaaccgatc gtagtccgga tcgcagtcgt caactcgact gcgtgaagtc ggaatcgcta 1260
gtaatcgca atcagaatg t
1281

```



<210> 107

<211> 43

<212> ADN

<213> Séquence artificielle

<220>

<223> Description de la séquence artificielle:amorce

<400> 107

cgctgcagat ttaaatatgc aacgcgtaag tcgatggcgt tcg 43

<210> 108

<211> 51

<212> ADN

<213> Séquence artificielle

<220>

<223> Description de la séquence artificielle:amorce

<400> 108

cggtcaactt aattaagata tctogagaga tctattaata cgatacctgc g 51

<210> 109

<211> 29

<212> ADN

<213> Séquence artificielle

<220>

<223> Description de la séquence artificielle:amorce

<400> 109

aaaaagatat ctgacgtccc gaagcgctg 29

<210> 110

<211> 32

<212> ADN

<213> Séquence artificielle

<220>

<223> Description de la séquence artificielle:amorce

<400> 110

aaaaaagatc tggctaacta actaaaccga ga 32

<210> 111

<211> 36

&lt;212&gt; ADN

&lt;213&gt; Séquence artificielle

&lt;220&gt;

&lt;223&gt; Description de la séquence artificielle:amorce

&lt;400&gt; 111

gtgccgttaa ttaagctccg cgaagtcgtc cttctt 36

&lt;210&gt; 112

&lt;211&gt; 36

&lt;212&gt; ADN

&lt;213&gt; Séquence artificielle

&lt;220&gt;

&lt;223&gt; Description de la séquence artificielle:amorce

&lt;400&gt; 112

gtgccgttaa ttaaccgctg cataaccctg cttcgg 36

&lt;210&gt; 113

&lt;211&gt; 42717

&lt;212&gt; ADN

&lt;213&gt; Séquence artificielle

&lt;220&gt;

<223> Description de la séquence artificielle:cosmide  
a26g1 brin non codant

&lt;400&gt; 113

```

aaaaaaaaa aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa aaaaaaaacc gggccctgac 60
gttcagaaact ccccgcgaga atctctcggc agagcgctcg cactctcgact tcaccggcag 120
tgtcgagatc gatgcagggtg caagcgaaact cgggatgctc ggcccgcatg gtacgtccca 180
gacccacacac cgggtccccgc gggggcacca ccggcatctg cagatgtctcc gcacgaacgc 240
ccgcccgtaat cagccagatt ccgggatgct gcgattcggc gtccgacgct tgtttcgtca 300
gctgtatgcy tccactccg aattcttgaa cgatgcgcaa aatgtcttcg caggcggttc 360
cccccgcagc cgacgggatcg gtctcatcct cggctctcct caggctcccg cagtgaatga 420
cgccccggta aggctgatcc ggtatctcgg cctcgcgacc cgaatcaca accgtgtttg 480
tgcccagccc tcgcccaccc cgggctcgca tgcgcggcgc atcgccaatg accagccatg 540
caccgcacac cgttgccgtt ggactctcgg ccagcagctg cggtccccat tccatagcgt 600
ggaaacattc ggattggcgc tccagttcct gggcacgcag gccttggacc tcgaggatga 660
cgtggccttc cgcatacacac agggtgacat cgcctctcag ccgccccgtc agcccgcat 720
gcaccctaag atcgccggcg ggtctgccga aacagtgcga ccgttcgatg gcgacaggca 780
cgcaaggacc ggcgctgcct tcgcccgaac gcgtcgcgcc cagcacctgc aaacaggcat 840
cgagcaaggc aggatgaagc gtgtaaccgg actctgtctc gcgaacggca tccggcacgc 900
tcagtcgcgc cactgcctcg ccgtcgcgcc gccacaactc ccgcatgccg cggaagggtg 960
cgccgtaatg catccctgc gatgcgaagg ccgcatagaa gtcctcgccc tcgatgcgat 1020
ccccaaagtgt gggcaggctc accgtgggcg cgacctgtgc cggcgccgca gccatggtgc 1080

```

cgcgcgcgtg ctccgtccaa tcggaaccgc cttcggccag actggagatg cggaatgcgt 1140  
 gtccctcgag tatgacctgc actcgcgagg cgccgcgga aggaacaacc agcatttgtt 1200  
 caaacccgat ctcttccagg ctgcagccac ccgcgaacac ttccctgggt gcggccagcg 1260  
 ccattctccac ataagcggca ccgggaagca ccacaagctc gttgagccgg ttgatcggcg 1320  
 gaaacggcga cgcattccaga gagagcacgg actcccgagc gtgtgtgtcc gcggccagcg 1380  
 caatctcgac cttcgcagcg agcagcggat gaccgcaca tgccggcaaa ctctccgcgcg 1440  
 tcgaggtcgc gaaccgaag ogctcgcgt gccagggata cgtcggcaga tccaggcgcg 1500  
 tgtcgggaga cgaagcgagc gcgcgccagt ccggacgctg cccattcacg tagagcgcgg 1560  
 cgagcaactc gagcagctca cgcgcgtccg gttcgtcgcg gcgcagtagc gggcgaaaca 1620  
 gtccgtttat gcgcagcgtc cgcagactat cctcgatcga cggcgtcagc acaggtcgcg 1680  
 gactgatctc caggaaactgc gtgaactcat caccagccat ccgctgcaac gactcccgga 1740  
 aacggagctg ctgtcgcaga ttggctaccc agtaacgcgc gtgcgacgcc tcgcccgctgc 1800  
 tcaactgtcc ttcaaccgtg gagaagaacg gcacggcgga acgttttgca ataacgcggc 1860  
 cgagttcctg gcgcaattcg ttctcagagc ggtccacctg cgagctgtgt gaagcgacat 1920  
 ccacctgtaat cagccggcag aagacgcgcg gcctctcgaa gtgcgtcttc aaatgctcga 1980  
 gagccacacg gtctcccgag aacaccgtgc tgctgtgtcc gttgtgtgcc gcgacagaaa 2040  
 cagtagtgag accgcgttca gcgagcacgg ccttcgcccc atcgagcggc agttcgacca 2100  
 gagcctcgcg tcgccggcgc cgaagtccga gcaacagccg gctgcggcga cagtagatgc 2160  
 gggccgcgtc ctccagggtg agaattcctg cgcacatggg ccgccccatc tctcccatgc 2220  
 tgtgtccggc cagcgcgtcc gggcgaaatc cccaggattg cagcagttcg accagcgcga 2280  
 ttgaaacggc gaacagcgca ggctgcacgc gatcgatcgt gctcagccat gctcccgact 2340  
 cgtcggcgag caggtccgca agcgcgcatt ccacgaagct gcgggaaggc gcgtcgcaac 2400  
 gtctcgatgc cgatcggagc acaggctcgt cggaaatacag cggataccgc atgcgcgggt 2460  
 actgtccgcc ctggccggaa aagatgaagg caggtttcgg acgaactccg ggatcggcga 2520  
 aaccgctggc gacgcgcgca ttgggtttcat tgccgcggaa ggctcgcagc aatttgattga 2580  
 actcggcgag ggatgagggc acaaacgctg ccgcatgttc gtatgtcagc cgcgtcaggc 2640  
 tggcggcgga acacagcgcg gagagcggag cgtgaaagcg cccatcgca taggcgcggc 2700  
 cgagatcgcg cagagcctgc ggaatggcgc cgaagaaggc taggaggtat tcgcggccgt 2760  
 cttcggcatg gagttgatcc gggaccggca cttcctcgcg gctgggcggc cgcccgaccc 2820  
 cccgcgcgcg gcttcctggc cggctgtccc cctcgaggag ggggacacta gcatagcct 2880  
 tagctccac attcccttgg gcctctaac togccttgac ctctcttgac gtcttcgct 2940  
 ctacattcgt cttgcctct acattcgtct togtttctac attcgtcttc gcttctcgca 3000  
 ggacgacgtg cgaatttggt ccaactgatc cgaacgagct caccgccga actcgggggg 3060  
 ggccgttgga gggccatggc gaacatcgcg ttgctatctt tagcggcagc taattccaga 3120  
 gtacgtgggg gttgggcgcg ttgaaatgca gatggggcgg aatctctcgg tgctcgagg 3180  
 cgagaattgg cttgatcagg ccggcgatac ctgcgcgcgc ctcagggtg ccgaagtgg 3240  
 ttttcccgca cccgacgate aacgggaaat ccgcggcagc cccctgcgcc agaccgcgtg 3300  
 ccacgcgcgc cagttcgatg ggaatcccca gcggcgtccc ggttcctggg gcttccacgt 3360  
 aatcgacatc ggcgggggcc atgcggcgt ctttgagcgc cgcccgcaat acggcttcc 3420  
 gcgcgggacc gttcggcgcc gtgaggccgt tgctcggccc gcgctggtg acggccgac 3480  
 ccggaatcac cgcggaata cgateccgt caccgctcgc atcggaacgc cgttccaga 3540  
 cagcatccc gcatccctgc ccgcggcgt aacgctcggc ggagcgaggc aaactttgc 3600  
 aacggccatc ggccgcgatc gcccgaggc ggcaagaata gatcgtcgt tccggcgcca 3660  
 gaatcaggtt cagcgcgcgc gccagcgcca tgctgcactc tcgcgactgc aagctgcggc 3720  
 acgcagagtg aaccgcacg agtgaggagg gtgcagcggg aagctcgctc 3780  
 cctgcaacc cagcagatag gagatccgtc cggcggcagt gctgaacgcg gttccggatc 3840  
 cggtataggc gtcaatgagc gccggatcgg taggtttcag cggcgttag tcgtcgggtc 3900  
 tgatcccgat gaacactccg gtgtcgtcgc ccgcgagact gtcgggcggc cgacccgcac 3960  
 gctccaaagc ttcccatgcc acctcgagca gcaggcgtg ctcgggatcc agaccggcga 4020  
 cctcgcgcgg cgtgattccg aagaagccgg cgtcgaagcc gtcgacggca ccatcgagga 4080

atccgcccag acgcggtgtac atctttcccg gcgcgttggg atcgggatcg taaaacgcat 4140  
 cggcatccca acggcccgca ggaatttcgc ggatggcatc gatgccatcg tgcaggagct 4200  
 gccaaaatgc ttccggcgag ttccgcgcgg gaaacgcgga agccatgcgc acgatcgcca 4260  
 tgggttcggt gtggacgcgc tccagttcgt cgagacgcgc tcgcgtgcgc ttgagcgcca 4320  
 ggacgcgctg ttgaagagga ttgagatcgc tcctccagat ggtgcaaggt 4380  
 ttccggtgat aacgcccga tcctcctgct ggacagctcg cgcacttcac cgcacgagga 4440  
 atcgctgggg acgtcgagtc caggttcgcg gaggacatgg ccggccaatt ctctcagcgt 4500  
 cggatggtcg tatagcaatg tcgcgggaag gctcttgccg accagctctc cgatggcgcg 4560  
 cgccagatcc agcgcctatc gcgaatcgag tccgtattcc ttgagcggcc ggcgcgggct 4620  
 gagcgctctg gacgcacatc gccccagcac gccgcgggcc tgcttcgga tgcgcatctg 4680  
 cagcagttcc tgcgcgcctc ccggcgccgc ttccgtgagt tgctggatga agccgggatt 4740  
 gctcgcgggg ctccgccttt ttccggcgga gacttggaa acccgcatct gagcggcagt 4800  
 ctcccccagg agatcgccga agatgcgcgc acccaacttc gcggcgacga cgggtacccc 4860  
 cggcaggcct tgcgcgcgca tgccgcggcg catgccttcg cccgcccatt tgccccaaat 4920  
 gatgtccatc gcccgtagtc ctctgcgcgc gcgcgatgtg gcaaggctgt cgagaatgc 4980  
 gttggcgcgc gactaatgct tctgtccggc ggaaccgagc agcgaagcgg cgggaagaga 5040  
 gactacgaaa aagtcgagcg catggtggcg agtgagctgg tgaagattcc aggcaccctg 5100  
 cagcttcggg gccagcacct tctcgaaaac agccacacgc tgttctgtaa ctaccgcgtc 5160  
 atcgagcacg cctgcggcat gcacgactcc acgcagcggc tgggtgcgcg ggtcccgccag 5220  
 cagcgcgcgc agctgtttgct cggaaactcac atcgcaagca gcaacctga ctgcagcccc 5280  
 gactgtgctg agatcgccaa ctgcctccgt atgcggcgcc accagttacca gacggcgccg 5340  
 gccttgctcg atcaagcggc gtgccaccct tctctctaata gcgcgcgagc gcgcggtagt 5400  
 cagatcgcac ccgtcggtcg aaatggcagg cggcgccctc gacgtttcct tgtgccgcac 5460  
 cagcggcgga acgtagcggc gtcogtgcgc caatgcgcat gctttgtcgt cgcgcggcata 5520  
 accgatttca tccagcagca tggcggcgcc gatgtcgcca ttgtcgcaac cgaggtcgat 5580  
 cagctcgccc gccagctcgg gatgctcgcc cgcgcatgcc tgcccagctg cccacagctg 5640  
 agcctggaaa ggcacgacgg gactgcgcat gtcactactg atcgatgca cgcgcgcgt 5700  
 catcagccag agccgcgcgc ggcggcgcgc caaagtctgg gtctgttcca gcgcctggcg 5760  
 gagagtgcg ccggacacga cgcgcagctc ttgcggcatc aaaccggcgca catcgtctgc 5820  
 gccggcacac agcaaccagt cctccggctt gccaggggcg tcggacttca acgttgtcga 5880  
 tcgctcgcac tcttgccact gcacatcctg cagccacgac tgttcggatt gcaagctacc 5940  
 ggcatgcatt acagccagtc cggaaaaactc ggcgatgacc gcgcgcgtct ctccaaccag 6000  
 cgtcatgata ccgacgaacg ggccgcgtcga gctcggggcg agacgcgcat gacagcgcac 6060  
 agaacctcgc gccggacggt agaagcgcac cgtctcgat ccgacggcgca cgtatcgccc 6120  
 gggctggcaa cgtccgcggc gccaaagtgc tcggaatact tgaaaaacag aatcgatcac 6180  
 gccgggtgac agccggtaa cgttcgcgcc atctcagcc accggcagac gcattcgccc 6240  
 cagcgcctcg ccatcgcgac gccagacttc ttccaccocaa ctgaaggcgg ggccaagatc 6300  
 gacgcgcggt gcgttcatcg ccgcgtagaa cgcacgcgcg gaaatgacct cgggaaggctg 6360  
 cgccgcgagc tcgaaatgaa cggcgcgcgc agtcgcgcgc cgcagactgg ctctcggtg 6420  
 gagcttccac gaatcgccat cctggctgaa gacctgacc tttgcttcgc cgtcctcgcc 6480  
 ggggtgtgata atcgcttga cctgtgacgg cgtatccggc gggatggcca gtgcctcgcc 6540  
 catcatgaca tcggagtcgg cgcagggaa cgtctcgat ccgacccgaag acttctcgat 6600  
 aaatgccgac acgtgcgagg ccgcggggcac aatgaccgcg tcgtagatca cgtgctcatg 6660  
 gagcagagcg gctcctggtg ttagcgaatt ttogaagatg acatcgccca acgcgctgtt 6720  
 gaggcgcgct cccaacatgc gcgcgcgcgc cggctctctc gcgggtacgc gctcaggct 6780  
 gaaggtgtca cgttgaacgc gatacgtcgg cagcgcgagc cggctgggtg attccccgc 6840  
 atagagacgc ccgactcgcg gattcacgcg ccggttaaac aggcgcgcaa gactttccag 6900  
 cagcacggac caatccgatg gtcccttaga tagggagtcg agccagacgc cgcgctcatc 6960  
 gggcagacaa tatcgcccca gcgtggtgag cgtgggatgc gggcgatatt ccagaacacg 7020  
 cttgcactcg cgttcgcga cgggttcgat cgcgctttca aactgcacgc tttcgcgcaa 7080

|             |             |             |             |             |             |       |
|-------------|-------------|-------------|-------------|-------------|-------------|-------|
| ctgtcgccgc  | cagtagcggg  | cgctcgatgt  | cgctgccttc  | ggcaataacgg | ctccgctgac  | 7140  |
| gttcgacacc  | agcgggateg  | ccagcggctg  | atacgcgata  | gcacctcgaa  | gcgcttcgaa  | 7200  |
| cttgtccaaa  | atcggatcca  | tcagcggcga  | atggaaacga  | tgcgatacgt  | tcagctctcgc | 7260  |
| cgtttccacg  | ccggcgcgat  | gcaggctate  | ttgcgcttcc  | gcgattttctg | cagccgtgcc  | 7320  |
| ggagatcacg  | gtgcggcttcg | gcgcattcgc  | tgcggcgact  | gccaccttgg  | cggcgagcgc  | 7380  |
| cgcgatcgcg  | ctcggaattgg | cgtagaacgat | gaccgctttg  | ccgcggggaa  | gcgcattgac  | 7440  |
| cagccgcccc  | ctggcggtca  | ccagccgcag  | gccgtctctcc | acgctaaaag  | ggccggccac  | 7500  |
| acacgcgcga  | acatactcgc  | cgagactgtg  | gcccagcacg  | tagtccggcc  | gcgacgcgag  | 7560  |
| cgacagccag  | aactgagcca  | gagcccatct  | caggggcaaac | attgccggat  | gggtatacgc  | 7620  |
| cgctctcgtc  | aacggcgaaa  | ccgactcgaa  | caagagaaacg | gtcagcgga   | catcgagctg  | 7680  |
| gggacggagc  | caatcggcgc  | aacgatcgag  | cgctcgcgga  | aaaaacaggct | gcgttttata  | 7740  |
| aaagctcgcg  | cccattgcgc  | cgtactgcgc  | gccttggcgc  | gtgaagagaa  | aagcgattgc  | 7800  |
| cggcgccggg  | cgcaacgata  | cttcgcgcgc  | cggcgccgcg  | gccaatgccg  | ctacagcctc  | 7860  |
| tgccgcgatc  | gcggcggtga  | tcgccaagcg  | gtgactata   | gcgtcgcgact | caacctgagc  | 7920  |
| ggtagagcaa  | acgtcggaca  | gcaacgcatt  | cggtgcgac   | tgcgaagact  | cccgcaagtgc | 7980  |
| gcgggacagt  | tcggcgagcg  | cttcgctcgt  | gcgcgccgac  | agagtgcgaa  | gctgcgggcg  | 8040  |
| tgtagaccgc  | ttcggcaaaag | ggagtgccag  | cgctctcttc  | aggatgacgt  | gcgcgtttgc  | 8100  |
| cccttccaaa  | cgcaacgcgc  | tgcgcgcgcg  | cagacgcggc  | cgctcttcgc  | acgtccaagg  | 8160  |
| cgacgatctc  | gtggcgatgc  | gaaaccggct  | gccgtccagt  | gagatgttcg  | gattcagccg  | 8220  |
| gcgaaaaatgc | aggtgcggag  | gaatggtgcg  | atgctgcagg  | gcgagtaacg  | ctttgatcac  | 8280  |
| cccggcgatt  | cccgcgcgcg  | ccctccagatg | cccgatgttg  | gtctttacgg  | aacccaagac  | 8340  |
| acaaggcgca  | gagtcggcgc  | cgctcgtagc  | cgactgcagg  | gcctcgatct  | cgatggagtc  | 8400  |
| gccacgcgcg  | gtcccgctgc  | catcgccctc  | gatcaacgat  | acgtgggagt  | gatcgatgtg  | 8460  |
| cgctgtggcc  | acgcctctct  | gcaggaccgc  | cttctgcgcc  | tgcagattcg  | gcgcgctgat  | 8520  |
| gccattgtct  | cgctcgctct  | gattgattgc  | cgagccgcgc  | atgatctcac  | ggatgcgctc  | 8580  |
| gccactggcc  | cgctcgctcg  | agagccgctt  | cagcagcacg  | atgccgcagc  | cttcgcgcgc  | 8640  |
| cacataaccg  | tcggctgcgc  | cgtagaacgt  | cttgacgcgt  | ccgtcgggcg  | ccaacatgcg  | 8700  |
| agccttcgac  | aaagcgatac  | tgccctcggg  | agtcaggatc  | aagttcactc  | cgccggcgaa  | 8760  |
| tgccgcgatc  | cattcgcgcc  | ggcgaggctt  | ttggcaagcc  | agatggacgc  | cgacgagcgc  | 8820  |
| ggaggagcag  | gccgatcgca  | ccgccatgct  | cggaccgcgc  | aggtcgagca  | gataggagat  | 8880  |
| gcgattggcc  | aacatctgat  | gcgccacgcg  | ggaaaccgcg  | caagctccga  | tgcgggcagg  | 8940  |
| gtcggcgatc  | tgaaacagtc  | cgaaagtcct  | ggcgaggagc  | ccggcaaaag  | cgccggtcgc  | 9000  |
| gctgcccgcc  | agagggccgg  | gagagatgcc  | ggcgctctct  | gccgcttccc  | agcacacttc  | 9060  |
| cagcagcagc  | cgctgctgcg  | gatccatgtt  | cagagcttcg  | cgggggggaga | tgcgaaagaa  | 9120  |
| ttccgcgatc  | aaaccgtcaa  | tcgcttcgag  | gaaggcgcca  | tatcgcgcgt  | acgccttgcc  | 9180  |
| cggagcatcg  | ggatcggagg  | agtagtactg  | gtccgagttc  | cagcggctcg  | gcggcacctc  | 9240  |
| ggtagacacc  | tgacacagct  | ttctcaaacg  | gcctccagag  | gcgtccggat  | ttctcgccgc  | 9300  |
| gcccggaaaa  | cgacacgcga  | tgccgcagat  | ggcgatgggt  | tcggcggtga  | ccaggtcgaa  | 9360  |
| cttcgcgatg  | tttttcgcga  | gttttcgcgc  | caatagcgcg  | agtttgaccg  | acgacatggg  | 9420  |
| cgcaattttt  | tcctcaacga  | ccttgctcct  | cggagcgagc  | ccacggctgc  | ttcttccgac  | 9480  |
| atgtcgtcca  | accgcgtgag  | tgcagtttct  | atggcgggcg  | tgtctctctc  | cgacgagccc  | 9540  |
| gcggcgctcg  | cttcgacagc  | cggcagctccc | atttgcgagc  | cgaggtgcgg  | ggcaagacgc  | 9600  |
| gccacgctgg  | gatgacccca  | aatcagggtc  | gcgggggagc  | tgaagaccag  | tgtgagttcg  | 9660  |
| agacgggtgc  | gaaactccag  | ggccatgagg  | gaatcgaagc  | cgagttctct  | cagcggggcg  | 9720  |
| aggggatcga  | tagtttgaga  | gtcgatgcgc  | agcacgcgcg  | ccagctgctg  | ctgtagatgt  | 9780  |
| ttcttcgagca | atgtctctcg  | ggctcgaggc  | tcggccgatt  | cgacgcgcgc  | gcgcaagcgc  | 9840  |
| tttggcgcat  | cggtctcgct  | cgccgcgtcg  | tcatgcgaaa  | gctcgaaacg  | tgcagatcgc  | 9900  |
| gcgcgcttgc  | gatagaactg  | ccgcactcgc  | cggacattga  | tgggcatcgc  | ggcgacgtgc  | 9960  |
| caagcgagcg  | tgttcagcag  | ctgttcacga  | atagcgagcg  | cggttgcggg  | cgtagggttt  | 10020 |
| tccatgcgcg  | gcgaagccag  | ccgcgatcgc  | cgaattgtcct | gcgcgggcgc  | cagcccgacc  | 10080 |

|             |             |            |             |             |             |       |
|-------------|-------------|------------|-------------|-------------|-------------|-------|
| tccgaccacg  | caccccaacc  | gatgctcagc | gccggcaggc  | cttgggcctt  | ccggtagtag  | 10140 |
| gccagcgcg   | caagaaagcc  | gttcgcggcc | gcgtagtttc  | cctggggcggg | cgcgcccaagc | 10200 |
| agtcctgcag  | cggaggagaa  | gagcacgaaa | tgatcgcagc  | ggcagctcgcg | ggtagagcaag | 10260 |
| tgcaggttcc  | aggcacgcgt  | gattttcgcg | gccatcacgt  | tcgcgaaatg  | cgcttccgcgc | 10320 |
| tggttcagta  | gcagcgcatc  | gtcgagaacg | gctgcggcat  | gaatcacgcc  | gcgcaatcga  | 10380 |
| tcgatggaag  | agatcacgcg  | ctcagattca | tcgcgctgag  | aaacatcggc  | ctgcaccgctc | 10440 |
| cggacatctg  | cgtccatgac  | ggcgatggct | tgctggacct  | cgggtgaagg  | cgcgcgccgg  | 10500 |
| ctcagcagca  | ccagcccgcc  | ggcgccgcgt | ccgatcatcc  | agcgtgcgac  | ggtaagaccg  | 10560 |
| agcccgcgca  | gtccgcgcgt  | aatcaagtag | gttccctcgc  | tatcgaacgc  | cgagcgtagg  | 10620 |
| ggtgcgatgg  | gcgcattggc  | gcaatctcgc | atcgccatga  | cgattttgcc  | gatgtgcgcg  | 10680 |
| gcctcgccca  | tggctgcgaa  | cgccctccac | gattctggta  | tggtcgtcac  | tcgcgtttccc | 10740 |
| agggggcgcc  | agggttccga  | ttcgaatttt | gcgaccatct  | cctgcagcag  | ctcccgggtgc | 10800 |
| aatgccgggc  | gcttcaggga  | catgccgagc | aaatcgacca  | gcgtgtacga  | gagggttcttc | 10860 |
| aggaacgggc  | gaagccccag  | cttgcggccg | gcatagtaat  | cgcgcttgcc  | gatctcgaag  | 10920 |
| aaccgccat   | gatcgcgcag  | cagatcgaag | ctcgccctcca | gcagatcgcc  | ggaaagcgaa  | 10980 |
| ttcaggacga  | cgtctactcc  | ttcttgattc | gtccaatgct  | ggatgtcgctc | cagaagaagcc | 11040 |
| atcgagcgcg  | aatccgaacc  | atcgcgcatg | cccagcgagc  | gcagatcacg  | tcgtttttcc  | 11100 |
| ggactcccg   | cagtagcgaa  | gatctcgcgc | cccgacgcct  | gtgcgatctg  | gattgcccgc  | 11160 |
| aatcccacac  | cgcccggtggc | agcgtgaatc | aggactcggt  | cgccggggcgc | cgccggggcgc | 11220 |
| gctcgcgaga  | gcgcgtaatc  | ggcggtgaga | aacgcgcatg  | gcaggggcggc | ggcctggtcg  | 11280 |
| cgcggaatgt  | tggccggcgt  | caaggcaacg | cggaaaggcg  | gcgtggcgac  | gaagcgaccg  | 11340 |
| aaactgcaag  | gcgcaaggcg  | cacgacttca | tctccgatgc  | gaaagtcggt  | gacgcctttc  | 11400 |
| cccattggcca | cgatacggcc  | cgagcattgc | ccgcccaggc  | gcggggtgcc  | ggcaatcgcg  | 11460 |
| cggcgcgcat  | cgtcgggcat  | aacgcgcgag | gcgagcagaa  | cgtcgaggaa  | gttcaggccc  | 11520 |
| gcggcgcgga  | cttcaatctc  | cacttcaccg | gcttgcgggg  | ggcgcgcgga  | tgtggcccgc  | 11580 |
| aagcgacgcc  | ggtcgaggac  | tcggggggca | tcgatctcga  | gccggaacgg  | ccgatcgccg  | 11640 |
| gccttgaaac  | tggcggggtg  | catatccgct | tcgtgcccag  | ccacgcgcgc  | gacgtaacgc  | 11700 |
| gcgcgcgcgc  | gaaaggcgat  | ttgattctcg | ccgttgttgc  | tcagcagttc  | gtgcaggagt  | 11760 |
| tcctctctgc  | cgcccgccgg  | atcgagatgc | atcacgctgc  | agttcagttc  | cagatgtctg  | 11820 |
| taatgcacgg  | tcggcgccca  | accccagaaa | ggcgcttgag  | cgataccggc  | ttgcaggatc  | 11880 |
| tgctccatga  | ccgggtgcgc  | gcgcgcgctg | accagccata  | ggcgcggtgc  | ttgcagccag  | 11940 |
| ggcggtcgcc  | ccaggtctcg  | gaggagatgc | agaaatgcgt  | tgatgcaggg  | ttcgtgtcgct | 12000 |
| agcaaaaaca  | cgatttccct  | gagcggcgcc | tggagttcat  | cgagcttttc  | cgcgagggtc  | 12060 |
| tcgtctcacg  | ggttcggcgt  | agcgcgacgc | catcggggtg  | gcgcgctatc  | cacagcgccg  | 12120 |
| acaatgagcc  | atgaccgcgc  | cgctcgcgcc | gccggcggtg  | ctgcagcgcc  | gtgcgagctga | 12180 |
| gcgaccacag  | gcagttcgct  | caaccagcgc | cgcatgtcga  | tgcgctccga  | cgcatccagc  | 12240 |
| cgtcgcagcc  | cgagaccctc  | gatcggggoc | accagttgtc  | cctctccgct  | cagcagcgac  | 12300 |
| agatcgcgga  | taggtccttc  | cagcccgcca | tgctccacac  | ccacggaacg  | tgccgggatgc | 12360 |
| agccacgcga  | tcgggtcgat  | gccggcgggc | agccagggtc  | caccggcggg  | accaaaccgc  | 12420 |
| gcggcgatga  | ttcgcagaca  | tgcatcgagg | aacgcggcgc  | cagtgccagc  | cgtttccgag  | 12480 |
| ctacgcagac  | ggccgatcgc  | ctcacctgga | caactccaga  | tctgctcgag  | cgcccggaag  | 12540 |
| acggggccat  | actcgacgcc  | gtgctccgcc | acttgacgcg  | acagctccgc  | cgccggcacc  | 12600 |
| actgtggggc  | agcgggcccgt | caccgtctcc | gcagaatccg  | gcgggacggt  | cgatgcatcc  | 12660 |
| gcaggcgctc  | gacgaatgtc  | cccggaaaga | tgacgagacc  | atgtcgatgc  | ctcgccgctg  | 12720 |
| gaaatcgcaa  | acgacgcctc  | ccgggtgcta | tcgacccgga  | tggcggcgga  | caacgtcatg  | 12780 |
| ctgcgctcgc  | gcggcacaa   | gagcatctgt | gtgaaagtca  | catgctccag  | cacgcacgga  | 12840 |
| ctttcaccca  | aggtctcgga  | agttccggcc | agagccatat  | cgagatacgc  | agtagccggc  | 12900 |
| aattgacgact | cgccctcgac  | gcgatggctc | gccagccaag  | gcagcggaag  | gagactgagt  | 12960 |
| tcgctctccc  | agaagaaagt  | gccgggttgc | gtcgaggctt  | cgacgcgctt  | tcccaacacg  | 13020 |
| ggattgcccc  | acgtgatcgc  | gtgtgcgcgc | ggggaagcgt  | cgagccagaa  | acgacgacgc  | 13080 |

tgccagggat accggggcgag gcgcacgcaa ttgcccgaag ggtacacggt ccgccatgog 13140  
 acagtgtgcc cagctctcata gagggcgccc agcgacgtga gcatggaaacc gcgttcgtcc 13200  
 tggctggcgc gcagagacgg aaccagcgcc gcattgccgc cgatggcggg cagcaggatg 13260  
 ggatgagggc tgatctcgag aaagacatcg tgcccgtctg cggcaagatg gcggatgccc 13320  
 tgccagaaca gaacgcgcga tcgcagattg cgagcccgat acgtgctgtc gaggctgggt 13380  
 gtctccagcg tcgcgcgggt caccgtggag taaaaaggtat tggctcgcgg ccgcggttga 13440  
 atcccgctga gcgaactcgag gaggttcgctg cacaatgggt ccacttgccg gctatgcgcg 13500  
 gcgaagtcca ctttcaccgg ccggcaagac acgcctcgcc gctccagcgt ccgcagcacc 13560  
 tcggccaggg ctctcgacttc accggagatg acggtggagt tgggtccgtt ccgacccgcg 13620  
 ggcatagtc gtctcgtgta agtcgacagc acggcctcac attccgcgag ccgcagctcc 13680  
 accatcgcca tcccgcccg gcccgtgatc cggctcaaca gccggctgog gctgcaaatg 13740  
 atccgcgcgc catctcgag cgtcagcgca ccgcgacat gagcggcggc gacctctccc 13800  
 atgctgtgcc gatgacggc atccggctcg attccccagg aacgccacaa tggcggcatg 13860  
 cgacactgca gcgcgaagag cgcaggctga atgacctga cgcggtcgag ctccgcagat 13920  
 tcttcttcca gcaccagtc cacataaggc cgcattggcg cctcgcagcg ttcccaagcc 13980  
 tcgcgaataa cgggttcgcg gtccatccag ctgcgcccc ttcgcatcca ttgcgattcc 14040  
 tgtcccgaag agacgaatac cgtcttcctg cgtgggaag ggatcgtgat ccctggaggc 14100  
 tgccgcgcca gtctctcagc cgttctgcgc gaacacgca cgcggcatcg gtgatcgctg 14160  
 cggcggaact cggcgtgta gcaaaagatca cgcaggctcg gtgctgcga cgtctgcagc 14220  
 aattcccgt atgcccgcgc caccogaec agttcgtccg caccatgcgc ggacagcgga 14280  
 agcacatata tcgcgtctgc aatgcccgta gttgcgcgag ttgcctgcagt cctgcgaatg 14340  
 tcgggaatgt ctgcagtgtc gggagtgct gcatgttcgg gagtgcctcg aatgtcggga 14400  
 gtgccccag tgtcccctc gcgcaggggg acagccgccc gcgcagcgcc gcgggggggt 14460  
 cgggaatgga acccgctcgc agcttcgct ctaccagtc gcgcgcgttc ttcaagaacg 14520  
 acatgcgcgt tcgtgcggga ccaaccaaac gcgtgagcg ccgcgaacct tcgtctcgaa 14580  
 ccgcggggcc acggcggaac ttctctcaca atgtcgagc acgttccctc atccgggata 14640  
 ttccgggttca gctgtctcac gtgtaagctc ggcggtatcg tctcgtgact caatgcgagc 14700  
 accgotttaa tcaatccgc tatgctcgtc gctccctcca ggtggccgat gttcgatttc 14760  
 agggaccgca ccgcgcacac atcccgga ggtcgggga ggccgacggt ttccgcgagc 14820  
 gccctcgatc cgatgggagtc gccgagcgga gtcccctgc catgggcttc gatgtaacgc 14880  
 atctcgatcg cgcgcagccc cgcattggcc aatgcgacc ggatgacgac ctgctgagac 14940  
 acgacatgg gagcggtag ccgcggcgag cggccatctc gattgaccgc ggagccgcgc 15000  
 accacggccc acaccggctc tcgcggcgcg agtgcatcgg acaggcgctt cagcaccacc 15060  
 acgcgcgagc ctctccgaa cagcatgcgc tcgcgcgcgc cgtcgaaggc gcggcgagca 15120  
 ccgctggggc aggcgggttc catcttcgag gtggcgtaaa taaactccgc cgagaagcgc 15180  
 agattcactc gcgcggccac gggcagcgta cactgcgcgc tgcgcaggct gggcagcgc 15240  
 agatgaacgc ccgcagcgga agacagcgag gcggtgctga gcgcgatcgt gggctcttcg 15300  
 aagtctcaga aataggaaag tcggccgggc atcacgctat gcgcgctgcc ggtgcgggta 15360  
 tacggatcga tgcgcgcgc atcgcggtc tgcattcaga aatagtcgtc gctttggctg 15420  
 tggatccgga cgaagacgcc cgtcgggctg ccggagagcc ctccatcgt ctgccccgca 15480  
 tctccagtg cctccacgcg cacttccaac agcagcgct gcctcggtac aatgctgaag 15540  
 gccctcgctg gcgaaatgcc gaaaaatcg ttgtcgaac catcgatgac ctccagctgc 15600  
 ccggcttga tcttcaccgg cgtggcgggg ttcaacgatt tcaggatgcg ccggaccgac 15660  
 tctcgtccc atgtccagg cggtaacctc ggaatagcat cgactccact gcgcaacatc 15720  
 tgccagaact catcgggccc atcgccgccc ggaacccgac cagcatcgcg 15780  
 atgggttcgc gcgcgtcgcg ttcggccgca tcgagacgtc gctgatgctc ctccagctgc 15840  
 aggtacgctc gctgcaacgg ctaagggttg gggaaatcgt cggatctaga actcactcg 15900  
 aggtcctcga aaaatgagcg aactctgttt tcaacaaagc ttcgatttct ttgtccccc 15960  
 acccgcgat ctgggttcgc acggcgtcga gatcgtctgc agcggcggga ctccggctct 16020  
 cgcccgcgcc ggtgccaagg tagcaacggg tagcaacggc agcaacggct gaaggttcag 16080

cattgccggc catgctttcc agcggcagcg cgagcttgtc ggcgagatgc tgcgccaggg 16140  
 cggagaatgt cgggtaacgc cagatcaggg tggcagaag ctgcagcgc agcccgctt 16200  
 ccagacggtt gcgaaactcg agggccatca acgaatcgaa tccgagatca cccagcgtcg 16260  
 ctctcgccgt gagtttgcgt ggatcgaagc gcagcagctg tccggcttcg tgcatacgca 16320  
 gcgtttccag ccgcgcgcgg cgctgcgcgc cggctggaac tgcaggagcg tgcgtgcgca 16380  
 tgtcggccgc cgggtttggg tccgcggcgg cgggtgcgat gcggggcagc agggacatcg 16440  
 atgcggccga cggatagtaa cggagccact gcgcgatata gaagttccat acagcgacgt 16500  
 gcggccgaat ctgcgtcaat gctttgtaga gcgcgcgcaa tccctgttgc ggttgaataa 16560  
 ccgagatgcc gcgcgcggcc agacggttct cgcggttcgc ctgtgcggcc aaaccaacct 16620  
 gtgtccacgg tccccacgcy atgctgacgg cgggaagacc ctgggcgcgg cgcgagatgag 16680  
 ccagcgcgtc gaaaaatgaa ttgcggcggc cgtagttgcc ctggccggga gatccactg 16740  
 tcgcgctgcg ggaagagaag agaacaaaat gatccagcg cgggcggcg gtgagttcgt 16800  
 gcaggttcca cgccgcggct actttcggag ccatggcgcc ttcgaagcgt tcggtcgtga 16860  
 gattgagcag catgccgtgc gccagcgtgc ctgccagatg gaacacgccc gccaacggcg 16920  
 gtgttcgcg atcgatgac gcgagcgcat ccgatagctg ctgcgcttcc cccacgtccg 16980  
 catggatgat cttgacgttg acaccttcca gttgtggcgg aggaagctcg ctgcgtccca 17040  
 gcagaacgag atggcgcgct ccggcgcgcg cgagccatcc cgccacctgc actccagtc 17100  
 cgccgtagcc gcccgtagtc agataggttg cgtcgccacg gaacggccaca tgcgttcgca 17160  
 atgggactct gtgaagactg aggcgcggcg cccataccgt gccttgcggg atcgcaactt 17220  
 gatccctgc gatattgcac agcatcagcg tcgcgagatg cccgcagtcg ttgctgtgcg 17280  
 catcgagatc cagcagcgct cagcgcagct cgggatgctc ataggcaatc gtccgcgcaa 17340  
 ttccgtgcag ccaggcttgt cgaatatcaa tatctttgtc gggattgaga acccgggcag 17400  
 atccgcgctc cagcagccac aggcgcggcg gctcaggcca gcccgcttgc acgatgctgc 17460  
 gcaatacggg aagcaggtcg tcgatgcgcg gcgagggaca gtacacaat tgcagggcac 17520  
 gcggaccgga gcacgtatcg gccgtgcgcg aggtttggcc gcgcttttgc agagtctcgc 17580  
 caatcgccgc ctgcgcgatc acgagccaa gttccgcagc attgcggcca ttggcatgcg 17640  
 ccgcgggaac cgacgcggtc cattgcaacc tccaggtggg aatctccgat tcgccgagct 17700  
 ggcgcgtctc ggcgactctc gactgcagcc ccaccaatc cgcaccaag ctgccggtgc 17760  
 cgggtgacgag acggacatcc accgtggaat ccggccgcaa gaccggtat cccagcacg 17820  
 ggccagtggg cacttcagcg agcgagaatc ggtccagacc tacgggcaca tgcacatctt 17880  
 tcaaatcgtc gtgatggagc agggccgcgg gcaactgcag acagcagtcg atcgctgcga 17940  
 ttccgctcag cggaaatgcc acgcgacaaa gcacctcacc gttgcgcgc cagatggggc 18000  
 cgatgggttcg gaaggtggga ccgaagtgat agccgcgatc ccacagtcgc gaatagaagg 18060  
 catcgctgtg gagctgcgcg gtgcagcggg cgcgaatcgc atccagatcg atgggtgcgc 18120  
 ttgaaatgcc cgctcgagc atgccttcgc tgtgcagctt ccaggaaatc tcgcggtcg 18180  
 agatcgggaa ggaagctccg ccgcctctct catgaaggg taccagtga acctgcctgc 18240  
 cagcatcggt tctccggcagc gtcagcgcgc ccgtcaatga cagctgttcg acatggtag 18300  
 gcccgccgcc gagacctggy cgcgcagcgg cgagcgccat tgcaggtgct cacgctccgc 18360  
 gagtccagc cacaatcgty agccggtgat ccgcgaaatc tttccctcc acagtggact 18420  
 cgaactgcac ctccggcagc ggcgaaggga tccgcgggcc agggcaagcc tgagactcga 18480  
 cctgcggcgg acggatatcg atccaataac gctcacgctg ccagggatag ttggggcagcc 18540  
 ggccagtttg gccgcgcttg ggataaatac gagaccagtc cggagtgact ccgttgact 18600  
 gcagcgtccc cagcgtccgg cgcagtgoga ggtttccgtc ttcactgcgc cgcacagagg 18660  
 cagcggaact cgtgcggca tctccgagcy tttcctggat cggctggacc aacaacgggt 18720  
 ggggactcag ttccagaaa acatcatgac caccgcgcgc ggcctgcgc agccgcgtcg 18780  
 acagcatcac ggggtgggca agattacgag cccagtaacy agaaacacag tcttcacgc 18840  
 taactcgtcg gccgttgacg gtggagtaca ttccaaaggc ggccggcgc ggcctgaagc 18900  
 ctccccacc gcccggaac gccgcgcaca cggagtcct catagtggtg tgcgaggcaa 18960  
 tgtccacttt cagcgcagcy cagaagacgt ctttcgcctc cagttccgcg agcagttcgc 19020  
 ccagagctcg gctgtgcgcc gacaggacgg tgctgcgcgg cgtgttgcg gcggcaatcg 19080



agaccgcgac cgagcgcgcc gcatggcgag cgatggccct gccagcgct aattccacga 19140  
 cagccatttc tccctggcgc cgtactccgg cgagcatccg gctgcgcagg caaatcacc 19200  
 gagcggtctc atcgagagtc agcgcacctg caatgtgcgc tgcgcgcgact tgcgccatgc 19260  
 tgtggcgcat cacggcgctc ggctcgattc cccaatggcg ccacagtcgc gccaaaggca 19320  
 ccccgatcgc gaacaggggc gggtgaatca cgtcgatgcg tgcagcgccg cctgcgaact 19380  
 ctctgcgcag cgacagctgc acgtaaggct gcatggcgcg gccgcactct tcatggtggcg 19440  
 caccgaacac cggttcagaa gccatcaggt ccgggcccat gccgggcaac tccgatcctt 19500  
 gtcccgccaa aacgaaaaag acttttcgct tctggccgcg cggcacaaaa cctgtggcg 19560  
 tatccgggtt cgggttgccc gccagaaaac tgtccagccc ggccatcaag tctcgcgct 19620  
 tctgcgcggt gaatgcgcgc cgggtgtcgt atgaagtgcg gcgagcgcac gccgtgtagc 19680  
 aggtgtcgcg ggggtgtcgt ttaccacagt cgcggtatgc gcgcgcgaga tcacgcagcg 19740  
 cctccggact gcgcgcgcat agcggaagca ggtacgggtg gggcgactcg gacgcggcct 19800  
 gttgcggcgc ctgctcgatg agcacgtgcg catctgacc gctcaagccg aacgagttga 19860  
 tgcggggcag gcgcgcgcgc cggggtgcaa ccggccaggg ggtgagccgt gccgggattt 19920  
 caggtcgatc gggtgtccaa tcatgtgcg ggctggcggt ggtcacctct agatggggcg 19980  
 gaatggcttc gttctgcagc atcagcgcca ccttgatcag tgcggccacg cccgctgccc 20040  
 cctcgagggt gccgaagttg gttctcacc acccgagctt cagcttgttt ccgttgtgtc 20100  
 gccccgctcc cagcgcggcc gccagggtcg cggcttcgat gggatcgccc agcgcgctgc 20160  
 cggttccgct cgctcgaca tagctcacat ccagcgtctg caagcgcgcg tctcccaacg 20220  
 cctggcgcat cacggcttc tgtgcgggccc cgttcggcgc cgtcagttca ttgctgctc 20280  
 cgtccgtggt gattggcgtg ccgcgaatca ccgccatcac cggatcgcca tgcgcgacgc 20340  
 cgtcgagagc tgcgttcagc acaaccacac cgcagccctc acccgcgagc tagccgtctg 20400  
 ctgcgcctc ctgacgctta cagcgacctg cggctgcgat cgcttcagc ttgcagaagt 20460  
 agatcgtccg atccggcgag agaatcagat tgacgcgcgc gccagcgcg aggtcgcttt 20520  
 cacctgagcg caggctctga caggcaaggt gcaccgcgac cagcgatgac gagcatcgcc 20580  
 tgtcgatcgc catgttcggg cctcgagccg cagagatgta cgagagacgc ccgcggccaa 20640  
 cgctggccgt attgcccgtg ccggtgtacg cgtcgatatg cgcacccccg ccgcgcattt 20700  
 gcaggttgtt ataatcggtg gaaaagatcc ccatgaagac gccggtcccg cccccccga 20760  
 gccggtcggg tggaaagccc gegtctcga tcgctccca ggtgacttcc agaagcagcc 20820  
 gctgctgtgg atccaggctg atcgctcgc gcggagcgat gccgaagac cgggctcaa 20880  
 aacgctcaac ctgatcgatg aagccgcgct acccggtgta cattcgccc gtcgcgcgg 20940  
 gatccggatc gtatgagcca tcatgtctcc agcggtcggg tggaaactca cgtaccgcgc 21000  
 tgcggccctc gcgcagcaac gaccaatagg catcgagatt ggtatgcgcg tggaaagcgc 21060  
 agcccgccgc gatgagggcg atgggctcgc tgcgcgcgct cccagctggt tgcgtgcgtt 21120  
 tctgcacctt ctgcagcgca atcacggcgc ggcaagctt gctgagatcg tctgacccgc 21180  
 tcatgtttat tgcgtctcca accactggct gacctgcgcg agcccggaat cgagcagcgc 21240  
 ttccagttct tcgcggcgca ggttctcaaa ctccggcgct tccaccgggt atgcttcggg 21300  
 tggaaatacc gcatggagca cgttaactgac gatcgcatcg agcgacggat agtcgaacag 21360  
 cagactcgcg gccaaaggct gccccagtga ttgggagagc gaattgcgaa gttctatggc 21420  
 cattagcgaa tcgagtccta gttaccocaa aggtctgctg ggtatcgagc gttgtggaagt 21480  
 cgcgatgcgc acaaagcgcg ccagtgaact cctgatgtgc gcaatgagga tggcttcgcg 21540  
 ctgcgggggt tgggtctcgt tcaagcgggt gcgcagttga ggtgaaggca gcgcggcggt 21600  
 acgcagcaac tcgcccgtta tcgagcccg cggtagcgcg gcaatctgaa tggggcatc 21660  
 atgcaggagc gccctgagaa tgtgtagacc ctgctccacg gagaggctcg ccacgcgcgc 21720  
 catcgactgg cttggtgcgc cggccattcc gactcccgag cagcgccccc agttaatgt 21780  
 ggtcgcggcg aaaccacgtc cgcgcgggtg atgcggcag gcatcgagaa cggcgtgtgc 21840  
 cgcggcgtag cctgctctcc cggcaggacc taagacgag gatgcgcgat aaaagagcat 21900  
 gaagaagtgc agcggcagat cgcgggtgtg atgatggag tgtacagcgc cttccgcctt 21960  
 cggcgccatg acgcttcgca tccgcgtcca gtctgattc agcagtaacg cgtcgtccag 22020  
 cacaccgcg ccatggatga ccgcgcgcag cggtagactt cgggtgtgga tgcggcgcat 22080

gagatccgog acctcttctt cccggtgac gtegacgctc tctgccgtcg caccaatctg 22140  
 ttgcagcacg cgtctgctgct cctcgttttg aggcggcgcg ccggccagca cgacgcaggt 22200  
 ggccgctgtc tccaccatcc atttccgacg tgtaagtccc agggctctcga gcccgccggt 22260  
 gatcaataaa gtccgcccgc aaaccagacg gactgccgtc cgcgcgctgg gtccggcggt 22320  
 cagtcgcgcc acgtagcgcc ggttgcttct ccacgcogac tgatcttcgc cgtcgaatac 22380  
 acgcattcgt cggcgccgcg cggccgcgca agcatcgcca tcgtcgggat cagatcgat 22440  
 gagccgcccc cacagatccg ggtctcgcg cgcgatcacc cggcgaagc ccagagcgc 22500  
 ggctcgtatg ggatttgtga ccgcactggg cgcctgcgcg ccggccgcta ccagccatag 22560  
 ccggcgaccg gacttcaggg acttcaccag ggccagagtg ctgcggcgag cgagctcata 22620  
 atcatcgaga ctgtacaggt tgacgatccc gcgcagtgca cgtccaccga ctaggagcat 22680  
 gtactcgccg gctggcgcca cggtaaccga catctcgccc tgagctgtga gcgcattcgc 22740  
 cagagcgccg gccgcgcgcg cactgtcgcg caggatcagc catgccccag gctgtagctg 22800  
 tcgcaaggcg tggcgagagc gttccgggccc cactcgacc tcatacaatt cgggcttcog 22860  
 ttccgagcgc tggcccccct cgcgagtgac gcgcgggaaa ctacgcccct gaagtctccc 22920  
 gagaacgcga cctcccgagt ccagcaactg cgcctcgccg gtaaaagcgt cggcggaagt 22980  
 ccgggagaatt tgcgcggccc cccatacgcc gccctccagg ctgccgtaaa aacaaacgcg 23040  
 atcgataccc agcggagcga atatcggtat ttccggcgcca tccgcaagcg cgggactcgc 23100  
 ccgcgctcta agcaattgca ggccggccttc gcaccaatca caacggggat tgacggcggt 23160  
 tgcggaatac attcggcgca cgccttctcg ttacccgaaa tgaatgcgct gtagcgcgcg 23220  
 gtactcgccg cccagttcta tctcgaggtg gcgcagcagc gaatagtagc tgcttccatc 23280  
 caccgcagggc cggcgcttcat cgaccagtg cggcagcgga gcgacaccag cgtggcgggc 23340  
 aatattgccc gcagcatgta agttccacga gccgtcgagc aagctgagta tgcggaaacga 23400  
 ggcatcgccg tcatcgctct gtgaaagcac gagctgaaca gcgtgtcg gcgtccgctga 23460  
 aaggatcaga gggtcgcgca agttcacgct ttccagcgtg tgcggcgccg gcgcaaacac 23520  
 ctccgcggcg gccctcgagc ccatggccag gaagtacagc gccggggcca ccaccgaacc 23580  
 gtaatatcgg tggcttgaga gttagagcgga agcgtcgat agtttcgact cgaagataac 23640  
 gtctgcacc cggtagcaga gccggcacc cagcagacca ctgcgaaccg ctacagagttc 23700  
 cggttctgaa ctccgctcga tccaatggcg cgtctctcga aaagtagag ccggcagggc 23760  
 gacacgcctt cgcgaatacg gacggtcgaa ctctcgcaa tcgatgtcga acccacctg 23820  
 atatagcgtc gccacactgc tgagaatcgt ctcccatca ctcgccgctt tacgcagcga 23880  
 cggcagccac tgcttgccgt cgtcgggcag gcacttttg cccatcgca gtagaaccg 23940  
 cttaggaccg atctcgagaa cacgctcgca gccctcgccc ttgagcgttt ggataccgct 24000  
 ggcgaaacgg acagggtttc gagcgtgac tcgccagtac agcggattcg ccagctgtcc 24060  
 ctcgcggccc agttttgccg tgaggttcga aaccaagcgc atcgaaagat tgccgcaacc 24120  
 gatcgccgcc gcccgcgctt gcaggtccgc cagaatcgga tccatgctcg agctgtgaaa 24180  
 ggccgcgcca accggcagca tctcgctttt gatgccctcc gcaactagag ttgcagcgc 24240  
 gctctcaata tctcgcggcg caccgcgaat caccgacctca gcgggtccgt tgatggccc 24300  
 aatggagacg cgcgaggtga tctcgtcggc acagcgctgc tcgcccgcgt tgaccgcagc 24360  
 catcgacact tccggcaggt tctgcatgag ccggcccggt tcggcaacta agccgagcgc 24420  
 atccggcagg ctgacggcgc cggcaataca cgcgcccgcg tattcgccga cgtctgtctc 24480  
 catcaccagg tctggcgcta caccaccagga cttccacaac tgcgccagat cccactgcaa 24540  
 agcaaacact gcggcgctcg ccgcggcggt cgcgtcgagc aacgctcat cggcccaacg 24600  
 cgccgcgaga tcgagcgttc cattcagcag agctcgcgat tcatccatgg cggcgcgaaa 24660  
 caccgctgc gactcgtaga actggcggcc catgcccgcg tattcgccac cttgcccggt 24720  
 gaaaagaaac gcaatcttgg ggcgctgctg ggcgatcgga acccgctcgt cctccgtcag 24780  
 tctgttgcga gctcgtcgc tcgacccggc cacaatcgag atacggtgag ggaagtgcac 24840  
 gcgccttga ttggccgtga atcgacatc gcgaacgac aaacggggct ggttgttccat 24900  
 atggccgcta tacgagcga ccagttcttc gagggccgcg tctgtattgg cggacagcca 24960  
 aagcacatgt cgggatcgtt cgggcgcagc tgcggccggc gtcacggcg cgcgttgcct 25020  
 cagaaatcag tgagcgttgg tgcgcggat cccgaacgca ctgactgcgc ctgctctcgg 25080

ggtctttccg gcggggccagt cgagcagccg cgtactcaca cgaacccggag tgttttcgaa 25140  
 atcaattcgc ggattcggac gctggaaatt caggctggga ggaatctggc cgcgatggac 25200  
 ggcaagcacc gtcttgatca gccccggccac accggccgcg acgtctatagat gaccgatgtt 25260  
 ggctctgaag gatccgatat acacatcgcc gcttccgttt ttccggaaagt tggcagcgat 25320  
 ggccgagcatc tccaccggat ccgcgagcgg cgtggctgtt ccgtggccct cgatgtagcc 25380  
 gatggactcc ggcttccagc ccgccatctc ttgagtgcgc cgaatcaate cgcctcgacc 25440  
 gtccaccctt ggagcgttaa accccatgcy ctccggccca tcaattataa tagccgctcc 25500  
 gcgaatgagc gcgtagatcg tgtcgccatc ggccagagcy cggctcaagc gcttgaggac 25560  
 gaccacaccc gcgcgcttgc ccggcacogt gccttgagcy gactcatcga aggcgcggca 25620  
 gcgcgcgtcg ggccagcaga tcatgcccggt ctggtgcagg taccocacgy accgcggaac 25680  
 attgatggca actccccccg ccaaggcaat gtccgagcgg ccgogctgca agctctcgca 25740  
 tgccatcacc accgacacca gcgaggtgga gcacgcogtc tgaaccgtca ggctggggcc 25800  
 gcggagggttc agcttgtaa agacacgcgt ggccaggaaa tcttctgtct tggccgtcag 25860  
 cagctggtagc gcggaggggc gtgagaaatc gaacggctcc gcgggtggcg agtctgtcag 25920  
 caggtaggta ttgacgcgc atccccgcga aacgccgatc gaacctctat gctctcgccg 25980  
 cgcataatccc gcgtttctcca tgccttccca cgcgcactcg agaaacacgc gatgctcggy 26040  
 gtccatgatc tccgtctcgc cgggactgta gccgaagaac gcggcatcga aaaactcgat 26100  
 gccgtccagc agacccttgg ccgcgacgta gctcgggtcc tggaaagact ccgggctgat 26160  
 gccgcgccgc agcagatctt ccggcgaaag cctggcgatg gaatccacac cgtcgcgag 26220  
 attgcgcgag aactctctca cattgcgcgc ccccgggaaac cggccggcca tcccgataac 26280  
 tgcgatccga tcttctcgca ccgcagccgc aggttccgca cgcgcgggtc cggatttttc 26340  
 tgccagggcg gcaagcgact cgatcgtcgt atgcggaaac agatcgacga cggagagcgt 26400  
 caaccggag cgtctctcga gcagtccgc caacctgtg agcattagcy agtgcggcc 26460  
 gacatcaaa aagttctgcc gatagtcag gtgctccac cgcagaactt acgcccagat 26520  
 ggagcgcaat gtctccacca catcgccgc catcggtcg cgcgcagca cggcgctgtg 26580  
 gggaacaacg ggaagcgcgt tccgctcgat ttgcccgtt ggcgtagcag gaagggagga 26640  
 caggctgaca aaccccgagg ggatcatgta atcggaagg ccgcttgcca gccacgaccg 26700  
 caaatcgctc tgcagatcgc gcaogtcgcc cgttgccgga acgagatagg ccatcagccg 26760  
 atcgtctctc acgaccgtaa tgcctgctt cagcgcaatg tgcgtctcga tccggcctc 26820  
 aatctcgccc gggttcgatgc gaaacccgcg cagcttgatc tggcgatcga ctctgcccag 26880  
 gcactcgact gcgcgcgt cgg aacggtagc agccagatcg ccggtagagt aaatgcgtc 26940  
 tgcatacgc cactcgcgga atttctcac cgtgagctcg gggttcgat gatagccccg 27000  
 cgccagtcct gctctctcga tgtacagctc tcccgaaat ccggggggaa cgggctccat 27060  
 gcgcgaatcc aggatctata actcgtgtt gtcgatggga tggcgatcgc gcacgatcgt 27120  
 atcggaggca cccagctctt gtgtcttgt cagggccgac catatggtg tctccgtcgy 27180  
 tccgtaaaga tccacagct ctacgccact atcgaagt cggcgcccca gttccggcgy 27240  
 cagagcttca ccgcccgcga aaacacggaa gcctttacc gcttccagc cgaatccag 27300  
 caattgcgc caaccgctcg gggtcgcctg catgaccgta gcgccgact tatccagcag 27360  
 ggtgcgcgt cgtcgccgt caaacacgat ctgcgggtg cgcacgatga cgcggcgcc 27420  
 ggtgatcaac ggcagccaga tctccagctc ggcaatatc aatgacacgy tggtagcggc 27480  
 gaccagccca tggcgggctc tcagacccgg ctccgctcgc atggagcga gcatgatgac 27540  
 tagcgacgag tggcgatct ccaogccctt cgttcgcccc gtgcgcccc aggtatata 27600  
 gatgtaggcg agatcgtcgy gcttgctgcc gctgacgaga ttccagctt ctggttcgac 27660  
 ggagacogcc atcatcgca tcatcgccat catctcagcc accgctcct cgttgaggac 27720  
 cgcgtccggt tgcacttcat cgcgaatccg gcgagacga tcttgagct gcgcgggatc 27780  
 gagagctcgt tacycgtcgc cggacttcag aatcgcaag agcgcaatca ccatctccag 27840  
 cgagccctcc atcgccagag cgtatgactt tcccgggccc gcgcggatg cgtcagacg 27900  
 atgagccagg cggttggccc gcgcattcag ctccgctgag gtcaactgat ggtcttcgaa 27960  
 gacaacggcg acggcgtcgy gagtgcgtt cgcctgagct tcgaccagtt catcgacga 28020  
 cccgttcgga ccggcatcgc ccggtgtcgc attgtgtcgc tcgagcatcc gcttcggacg 28080

cgcggggggac aacagcgcag cgggttgaat gcggacgtcg ggatccgtca ccacgtctcg 28140  
 cagcagggtt cggtagcat cgaacagggg ggcgatgggt gccgcatcga acaaatcgg 28200  
 gttgtattcg gcggacgcca tcagtccatc gccgatggc tcgaggggtca cgcggagggtc 28260  
 gagtttggat ccgcggttgt gcatgtactc gcgcgagatg gtgagccag gcatgacggt 28320  
 gatggccggc gcatcgggca gcagcgcgaa ggagacctga aatacaggcg accggtctag 28380  
 gtcccgcgga ggatgcagtt cctcaaccag cgttcgaaa gaaagtctct gatgagagag 28440  
 ggcgctcaaa gcggtgtcgc ggtgtcgggc gagaagactg cgaacgacg gatcgtccgc 28500  
 cagatgcgcc gcgaggagca tcatgttggc gaaacaaccg acgagacctt cgtttctcgc 28560  
 ttgtgtacgg cccgcgactg gaaccccgat aaggatgtct tctcgcgcgg tatagcgatg 28620  
 cagcagcacc tgaaacgcgc cgattgcctg catgaacacc ctgcctcctt cagcaaggc 28680  
 aaacgcgtgg agtccatcgc tcaaatcacg gccaggggct gtggtctcca cggcgcgccg 28740  
 ccagggtctg tgcgcggggc gggggcgatc ggtaggaaag tcgaggaaag gcaagggtgc 28800  
 cgacagctgt ttcttccagt actgctgcgc ggtttgggtc agcgacgtct gctgatggac 28860  
 ggcccagtcg ccatactgaa tcggcagttc catgagcggc gatggccgcc cctgcacgaa 28920  
 cgcttcgtac gatcgcgtca ggtcgcggac gaacgtctcg accgaccag catccgcgat 28980  
 gatgtggctc aacgtcagca gcggaatctg ctgcttgtca tcgaggcaga tcagcttggt 29040  
 ccgcagaaag gggggtttct gcagggtcgaa cgggatctgg gcatcacgca aggccatttg 29100  
 ccgcgcttcg cgcattccgt cagcctgaac aaccggaaat tccagtgta ctcgcccagc 29160  
 gaggtctctg cgcgctctct catccacacc gccaatgacg ctgcgcaggc tctcgtgcgc 29220  
 ctgcaccagc gctccagac tccgcaggag gacgcgaata tccagcggac ctcggatatg 29280  
 cagcgcctatg ggaatgttgt aggcgggaga atccgggtcg agctgatgga gaaaccaaag 29340  
 ccgctgctgg gccacgcaga aggggtcggc atcccggtt tcacgccgcg ggatgcatg 29400  
 ttccgggctg tttctcgcga gcagacggtc gagcaattgg cggcgggcga gcgagaggtc 29460  
 tatggatttt ggcagcgaat tctgcattac aaaccgcgtg gttcctagtc ttgggcggcg 29520  
 ctcatcatac cgtcgatttg aacatctgac atttgggaaa cagcgatcag caaatcggcg 29580  
 gctcgctctg cccatcctgg gtctacgcgc tctgaatag cgacggcgaa gccttcgcaac 29640  
 gtgggggctg taacacgggt tcgaaagggc acttccacgt ggagcatgtc gcgcacgcgc 29700  
 gcgatcatct cgtgaccag cagcgaatgt cctccagagt cgaagaagtg atcatggacg 29760  
 ccgatgccat ccattgcgag caccctgcgc caaatgtggg cgagtacctg ttccaccgga 29820  
 gtttccggag cgttgaatgc ttccgcgtgg gctcgcgcgc tgggctcggg atcgggcagg 29880  
 gcgttaccgt ccgattttct gttgggcgtc agcggcattt cgtggagcac gaccacgcg 29940  
 gtcgggatca ttagtcggg cagcttctcc ttcagatgag tacgcaactg gcgcacaacc 30000  
 gtgcgcgatg agacggctcg cagcggatcg ttctgatcag ggcgcggcag cgcgcgtcgc 30060  
 ggacgggaaag ccggcggaacc gcgcgacgca cgcgcgcaaa gcttcggtgt 30120  
 ggcccatgac tgctccagtc gattgccacg cgttacggca ggtcttcgtc catacgcct 30180  
 agatcggcgc gatcgacgcc ggaaggcgac gtctggcgca gcggttccgc caactccccg 30240  
 agtgtctctg gagcttcgtc accgttcacaa tggcgctttc ggcgggtcaac 30300  
 cgtgcgtctg gaattcgtgt aaatgcggcc aactccgctg gacgctcgtc cagtactctg 30360  
 cgtatttcgc ccgcgggtcg gcaacgcctg cgaatcgatt ccggctctct cgttccccgc 30420  
 gatccgatat gcaggatcgc ctggtagcgg aagcgggtca gctcgttatg cgacgcgcgc 30480  
 cgacgcggca ggaattcaat ccggccgatc tccggaatct gtcgcgggag agcaagaag 30540  
 aacgcgggat gcaccagag ttctcttctc tcgcacgcga gcgaacgcac gcttgcgca 30600  
 aactcattcc ggggtcaacga cgcgggtgcg cgtgaactt ctaagaagc gtaaaacgtc 30660  
 tccagcagcg ggagactgcg gacatcgccg acaaatacga tgccgcggcg tttgaccaca 30720  
 cgacgcgctc cggccagcac gcgcgcgaga tacgcttcgc cgggggaagta ctggaatacg 30780  
 gagttcagaa caacgcgcat gcacgagcga ctgtcgatct cgcacgcgtc gtcggccgc 30840  
 tgcgggaacg tgcggaact ttgcaggccg gtcgggtccg cgtgagcggc gatgtagtc 30900  
 agcgccttct gcgaaagtc cgtggccgag tactcgaag agcggggaag gacggggaag 30960  
 agcagcagtc ccgtaccaca gccaatctcg agcagcgcag gcggccgcga ggccaggatg 31020  
 cgatcgacgg aatcctgcac ccaatccgc atctcggcag ctggaatcgc ctctccggt 31080

acactgcttc tccagccgac gatgttgaac tccggatccg cgttcggcgc attctgttca 31140  
 tatgtggtgt cccagacgga ttgccactgc gtacactgac cggactcgac tcggtcggtg 31200  
 aatgtgtcgg cggctgcgct cgccgcatgc cgttcagcaa gggggacaat gtaggccgcc 31260  
 agatacttac cggccgcgct attttctctg gcggtgacca cagcatgtcg gaccgcgggg 31320  
 tgactcgcca ccgcccgcct gatctcgccg gtttcgatgc ggaacccgcg tatcttcacc 31380  
 tgggtggtca tccggccgag atactcgagc gcgcgctcgc gttggccggc ggcgagatct 31440  
 ccgtcgcat acagccgagt gccatgaggg tcgaacgaat ttggcgacga cctgcgctcg 31500  
 ctgagttccg gacgattcag gtatccacgg gcgagcccg cgcgcgcgat gtacagttcg 31560  
 cccgaacac cgatgggtgc gggctgcac cagatcgtaa gcacatagag ctgagtggtt 31620  
 gcgatggggc ggccaatcga aaccgggtccg tcacctgtcg tcaccggttg gatggcgagc 31680  
 caaattgtcg ttctcgtagg tccgtaaaaga ttccatagcg ccgcggttcg ttgcaggagc 31740  
 cggtcggcaa gatcgcgagg aagggtctca ccgcgcgaga gcgcgtcag gcggcggtcg 31800  
 ccggggccagc cggatcgag cagcagacgc cagggtggcg gagttgcctg catcatgttc 31860  
 gctttgtctg gcgcgagttc cctcgccagc ctctcaccat cgacggccgt cctcgtggtc 31920  
 gccaccagca gcgcgcgcgc ggcgctcaag ggcaaaaaa tctcgagcgc ggaatgtcgt 31980  
 aacatgaacg tcgtgaggcg gacgagcgta tcgcggtcgc tgatgcccg cctatgccgc 32040  
 atcgacgaaa gaaaattgac gacggcctcg tgtgtgatt gacgccttt cggccggcgc 32100  
 gtcgaacccg aggtgtacag gacataggcg agatggcg cgggtcgag cgggttcgga 32160  
 ttggtgtcgg gctgcgtcca tacttccgat tccgtgacat tcgacacaa gaccggcgtc 32220  
 gtctcttcca gcatcagccg aagacgttgc gccggatatt ccggttccag cggaaactag 32280  
 gccgcgcggc ccttcaggac gcccaacagc cctgcgacgg ttccgagcga ccgcgtgaca 32340  
 tggatgccaa ccatttcgcc ggggtccagc ccgcgcgagc ggagatagtg cgcgatccgg 32400  
 ttggcgctcc cgttgagttc gcgatatgtc agattctgct caccgaagct caacgcgatg 32460  
 gcgtcggggc tcaactccac ctgagcttcg aacagctcgt gcacgatct ggacgggaat 32520  
 tccgcggcgg tcgcatcca ctcttcgagc agctggatgc gttccgggt tgtcagcagc 32580  
 ggtagatcga caactggaca ggcgggattc tccgcgattc cttccagcag caccggcgaag 32640  
 tgcaaggaga gacgttcaat cgtggcagca tcaaaaatgt ccgtgttgta ttgcagaaag 32700  
 gcggagagcg ctccatcggt ttccgaccat atcagatcca ggtcaaacgc gctctgtcgc 32760  
 agcggcatcg ccagggactc cagttgtgag ctgccccagg ccatgcgacc gccggactga 32820  
 cccaacatga accgacgga ttccgggaatg cgatgaggct gctggagcac gaatagaacc 32880  
 cgagtcggg gaccacacgc ctccacgata cgggcatacg ggtactcctg gtgctcgatc 32940  
 gcgcgcgaaa gcgtttcgcc aatccggggc agcacogtat tgaatccgg atccgctgaa 33000  
 agttctctc caggaattac gggattcacg aagtatccga cgagatcggc gaattccggt 33060  
 tgcgtccgac cgttggtgag ggtgcggctc aggatctctt cttgtgaggt ccaacgggag 33120  
 agaagcactt gaaacgcgcg catcagcgtc gcgatcgagc tcgcgtttcg ccgcgcgcgc 33180  
 atcgcttcca gtttcgagtg cagcgcgggt tgatttcgga acgagtaga gtttccccgg 33240  
 aaactctgca ccggcggact gggacgatcc gacgggagat tcgaaacccg aagctggccg 33300  
 gaaagctcg aggaccagta gttccaaagc cgctcgccct cggttccggc caacagtttc 33360  
 ttctccagc ggaacgaagc ggcggaagctc gcgcgacagc cggaccgcca 33420  
 gctgtcctcg caggttagat actcgggagt tcatccacca tcaccagcag tgaccagaag 33480  
 tcggcgagga gtgatgcac cacgatggcc agaactgat cctccccga ctgcaccagc 33540  
 agacgcgagc ggaacagatt ttgcgcgaga ttgaaggggc cgtggaagac gccgtcgatc 33600  
 agcacgcgct catcgtccgg cgaaacaggg atcacttcga aatccaccgc gacgtcgtc 33660  
 tggagcggtt gaacgggtgc gccgcactc tccgcaatcg cgttcgcag ccggcgatga 33720  
 cgatccacca ggtcctcgag cgaaacggcg aacgcctgcg gatcgaaagc gcctctcgcg 33780  
 gcgcgatcc accgatgtt gtatgcggga ctttccggcg cgcttcggta aataaaccaa 33840  
 agcgcgctgt gcggggcgct gagagggtag gagaggcgag gaaccgcgca ctcgcgcgca 33900  
 ggttccggcg ccacgctcgt cggttcgctg agggcgccta gatcgcttag atccctggcc 33960  
 agttccgcaa cgtcggggcg gttcagaat cggacccatg gcagcaagac gcgcagatcc 34020  
 gtatcgatcc ggttcgta ttgcaccgcc atgagcgagt ccaatcccat acgcaccagc 34080

ggctgctgta agtccacgcg cagattccggg cactgcagtt tttttttttt tttttttttt 34140  
 tttttttttt tttttttttt tttttttttt tttttttttt ttttaatgcg gtagttttatc 34200  
 acagtataat tgctaacgca gtccaggcacc gtgtatgaaa tctaacaaatg cgctcatcgt 34260  
 catcctcggc accgtcaccg tggatgctgt aggcataggc ttgggttatgc cggtactgac 34320  
 gggcctctgt cgggatgatc cctgtcagtc atgcggggcaa cttagccagag ccctacgaca 34380  
 ccgcccgtgg gaaggtgagt gtctaaactgc gtgacaacgc cagcgacagag cggcgacaaa 34440  
 ccgcgagcac ccatggactg gcgcgcgagg tgagaagcac actggcccaa ggtcgagcgc 34500  
 ccacccaagt tgcttcggga cgaagaggtc gtgggttcaa atccccccac ccgcagacag 34560  
 aaacaccagg tgaggcagac cgtaacgtta cgggtcgcct cactctgttt ctgtgcgtgt 34620  
 ctatctcggt gactatcgcg ccggaccgcc cttgaagatg ccgtccatga ccacagcgcc 34680  
 ggtctgcatg acgggcccga tctgcttccg gttagacctc tcaagtacag cgtaccgga 34740  
 gtgtccgagc agccgggaga tctcctccag cgggacgcgc cggtcggaca gacgggacac 34800  
 gaagctgtgc ctcaagctccc tgggtgtcca ctcgctggcg ttgatcccg tggcatcctt 34860  
 gagcgctcgt cggaaaggcg gccggacggt agtcgctgc agcgcgcttc caacggccga 34920  
 cgagaagacc aggcgctgtt cctcccaact gtcaacggcg gcgagcgctg cccagccctg 34980  
 gtctctaaag tgctgccaga ggacctccac gcaacgcgcc ggcaggggca gcttcgcgc 35040  
 agacttcggg gttttcgtgt cccacacgcgc ccggacgcag cgccacagcg cgaatgtgcg 35100  
 aggtctggcg ggctcaacgt ccggacttcc cttgaggaag acgtgtccc aggtcagcgc 35160  
 ccgcagctcc tgggtgcgcg caccggtcag caggggcagc acgatgtagg cgtgcacga 35220  
 cgtgcctcgc gcagcatcca gcaacgcctc ggctggggc aaggtgagcg ccttggacgg 35280  
 ccgcacagcg tggccctggg gcaacagagc cagctccacc agcttgcgt tcaacttgtc 35340  
 accgcgcatg gcccgcttga ccgcccgtt caggcaggag tggacgcct gaaggtctgc 35400  
 cgtgctcaga gtctagactc tggcgccag ccacgcgtgc agtctctctg cttgtaggtc 35460  
 accgagcttc cgggcaccca aaccgggtat gacgtgcttc tggcttaggt gggtagcagt 35520  
 ctcgacgggt cgtgtgtcac ggccagcgag accgtaggca agccagctgc tcaaccgcgc 35580  
 ggagctgggt taccocgtgg gtgcgatcgc gagacgctc tegtgtcac gcacacacctc 35640  
 tttgagcttg tctttagctc ccgtcttggt cttgccactc ccccgcttga cgaatcgctt 35700  
 accgctcga tcgaagcgca ggttcgcgt ggccatccag cgtgtctctc tctcgtccca 35760  
 gtggaggccg ccgtcacccc ggctacgtgc cttggccatg gatcgatccc ctgccggca 35820  
 aaatagagtg ttctctgtcc ctctttagca ttcagtgtat ccattaccgt catcaattgc 35880  
 tcactcccg ggccgctgtc gttgtcatcg aataaaatga gctgcgcgac tccttgactg 35940  
 aagaatcccc ccagcateac gcccgctttt tggtaacgat ggcccgcctg ccagatggca 36000  
 tccagagatc gcgtagcagc gttaatgata tcctgtgct cctgagtggt cgtcagcagt 36060  
 tttaccgagc cgtctattgc gtaataaggt tcatgtagcg caaatgggta cgtcttaata 36120  
 aacgtggaga taaaccgaca atattgatgc tcgctgcgaa gtttttcgcg cgcgggggca 36180  
 gcgtaactac aaatggcctg ccgcatcgac ggataatccg tgatgcgttc accaaaacag 36240  
 cgggaacaga taatttctct cctcgtcggt ccagttgcga acagggttcg 36300  
 ccgcgcagtt cagcgcacgt tctttcgagc acgacattaa aatggttacg gataaacggc 36360  
 atactgtat acgccaatc gagaacgggt ttgatccca tcgcgtccag tttttgtg 36420  
 atccgcgcgc caatcccca gacgtcatcc acggggagag cagacattaa tttacgctgg 36480  
 cggtccagat ttgataaact caccacccca ccgctctgcc gctgccattt ttttgcgca 36540  
 tgattgcgaa gctagcttta tgcttgtaaa ccggtttgtg aaaaaataa taaacttaata 36600  
 aaggggaact ctagggtccc caattaatga gtaataaat ctattaaagg tcaattcaaaa 36660  
 ggtcatccac cggatcagct tagtaaaagg ctgcgtagat ttaaatgcgg atgttgcgat 36720  
 tactttccca actatgcga taacaagaaa aagccagcgt ttcagtata atctcccaat 36780  
 ttgtgtaggc cttatctatgc acgcttaaaa ataataaaa cagactgtac ctgatagttt 36840  
 ggctcgtgag aattatgtgc ttagtgcac taacgcttga gttaaagcgc gccgcgaagc 36900  
 ggctcgtggt tgaacgaat gttagacatt atttgcgac taccttgggt atctcgctt 36960  
 tcacgtatgt gacaaattct tccaactgat ctgcgcgag cgaatcttgc cgaatcgga 37020  
 tggaaagccc gccgacccc cctggagag atgatcagg atgccagggc ctttaacgcc 37080

cgccgctgct gaggctccgc ccgcggggccc gcaccgcegt cggccggggccc gctccggggct 37140  
 cgcagcagcg ggcttcggcg cgggcccggg gctcccgggc cgcggggcg ggctccgccc 37200  
 ggcggcgccc gggggcgggg ggcggcgccc ggcggcccgg ggcgtcagcg gccggggggcg 37260  
 gtgtccggcg gccccagag gaactgccc agtctcccg gatccggtgaa gccggagaga 37320  
 tccagcgggg tctctcgaa caactcgaa tcgtgcagga aggtgaagcg gacgagttcg 37380  
 cgggcgaagt nctcggtcgc ctccactgc gccccgtcga gcagcggcg caggatctcg 37440  
 cggctcgccc ggaagcggtt gagatgcagt tgcaccagcg ttagtcggga gctcccgca 37500  
 tagacgtcgg tgaagtgcac gatcccgggt accctcggtc cggccaggtc cagcaagatg 37560  
 ttggctcccg gcaggtcgcc gtggacgaac cgggggttcg ggcggggcag cagcgtgtcc 37620  
 acgtccggca cagcgtcctc caggcggtcc agcagcggg gcgagaggtc gccccaccgc 37680  
 cgggtggtcct cgacgggtcgc cgcgcggcgt tcccgagca gttccgggaa gacctcgaa 37740  
 tgggggggtga gcacgggtgt cccgggtcag ggcaccctgt gcagcggccc gagcaccgg 37800  
 ccgaggttcgc gggccaggcg gagcagcgcg ttccggtcgg tcgtgccgtc catcgcgagc 37860  
 gccaggtggt tgccgggtcat ccgggtcatc accaggtagg gccacggcca ggctccgggt 37920  
 cggggccgca gcctgcggcg gccgaggagg cggggccacc gcaccggggc gtcccgagg 37980  
 accgogtaag cctccgactc cgacgcgagg ctctccggac cgcaccagtg ctcccgaaac 38040  
 agcttgatca cccggccggg ctcccgacc agtacggggt tgggtgctct gccgggcaac 38100  
 cgcagccagc cggcgaccgg cagcccgagc tctccaggcg ctccggcggc cagcggctcc 38160  
 cagaattcct ggtcgttcgc caggctcgc taggaatcat ccgaatacat acggtcgaga 38220  
 agtaacaggg attcttgtgt cacagcggac ctctattcac aggggtacgg cccggctaat 38280  
 tccgacggcg cggctcgac acggcctgtc cgcaccggcg atcaggctgt gacgatgacg 38340  
 gggtggtcgg ccacgtcggg gacgttctcg gtgggtgctg ggtcgggatc gccaatctct 38400  
 acggggcgac cgaggcgagc gtgtacgcca cagcttgccg taatcatggt atagctgtt 38460  
 tctgtgtga aattgttatt ccgctcacaat tccacacaa atacgagcgc gaagcataaa 38520  
 gtgtaaagct tgggggtgct aatgagtgag tcaactcaca ttacggatca gtaggggttt 38580  
 gcaactcgcg gtcacaggatc ttgatttcga tcaaggcagc atcatcgtgc gggaggcgaa 38640  
 gggctccaag gatcgggctc tgatgtttacc cgagagctgt gcaccagacc tgcgcgagca 38700  
 ggggaattga tccggtggat gaccttttga atgaccttta atagattata ttactaatta 38760  
 attggggacc cttagagggtcc ccttttttat tttaaaaatt ttttcacaaa acggtttaca 38820  
 agcataaagc tatcgtccat tccgacagca tcccgatca ctatggcgct ctgtagcgcg 38880  
 tatatcgct gatgcaattt ctatgcgcac ccgtttctcg agcactgtcc gacgcgtttg 38940  
 gccgcggccc agtctgctc gcttcgctac ttggagccac tatcgactac gcgatcatgg 39000  
 cgaccacacc cgtcctgtgg atctgcctcg ctggcctgcc gcagttcttc aacctccgg 39060  
 cgcagctttt cgtctcaca ttacgatcc ctctcgcat accattttat gacggcgcca 39120  
 gagtcaataa gcacctcatt acccttgcca ccgcctcgca gaacgggcat tccctgttcc 39180  
 tgccagttct gaatggtac gataactcgca ccgaaaatgt cagccagctg cttttttgtg 39240  
 acctccattg ttacttccac ggacaaaaac agagaaaagg aacacagcag gccaaaaagc 39300  
 tcgctttcag cacctgtcgt ttctcttctt ttccagaggt attttaataa aaaaacttaa 39360  
 gttatgacga agaagaacgg aaacgcctta aaccggaaaa tttcataaa tagcgaaaaac 39420  
 ccgcgaggtc gccgccccgt aacaaggcgg atcgccggaa aggacccgca aatgataata 39480  
 attatcaatt gcatactact gacggcactg ctgcagata acaccacgg ggaacattc 39540  
 catcatgat ccggtgcgga cataggaagc cagttctatc atcgctttct tgtctgtcgc 39600  
 catttgcttt gtgacatcca gcgcgcgaca ttacagcagc tttttcagcg cgttttcgat 39660  
 caacgnttca atgttggtat caacaccagg ttaactttg aacttatcgg cactgacggt 39720  
 tactttgttc tgcgctggct catcacgctg gataccaagg ctgattgtt agatatttgt 39780  
 caccggtcga ggtgtttcga ttgcgcctgc gtggatagca ccatttgcga tagcgcgctc 39840  
 ctgtgatgaat gacactccat tgcgaataag tgcgaaggag acggtgtcac gaatgcgctg 39900  
 gtccagctcg tcatgtcct tttgtgcagc agaggtatca atctcaacgc caagcgtcat 39960  
 cgaagcgcaa tatgtcgtc caccaaaagc cgtattgacc aggtgttcaa cgggaaattt 40020  
 ctgcctctct gatgtcagaa aggtaaaagt atttctttc tggattcag ttgctgtgtg 40080

```

tctggtttca gcaaaaccaa gctcgcgcaa ttcggctgtg ccagatttag aaggcagatc 40140
accagacagc aacgcgccac ggaaaaacag cgcatacaga acatccgtcg ccgcgccgga 40200
caacgtgata attttatgac ccatgattta ttctctttta gacgtgagcc tgcgcacag 40260
caaagccgcc gaaagttaac ttgtttattg cagcttataa tggttacaaa taagcaata 40320
gcatacaaaa ttccacaaat aaagcatttt ttctactgca ttctagtgtg ggtttgtcca 40380
aactcatcaa tgtatcttat catgtctgga tctgacgggt gcgcatgac gtgctcctgt 40440
cgttgaggac ccggctaggc tggcgggggt gccttactgg tttagcagaat gaatcaccca 40500
tacgcgagcg aacgtgaagc gactgctgct gcaaaacgtc tgcgacctga gcaacaacat 40560
gaatggtctt ccggtttccgt gtttcgtaaa gtctggaaac gcgggaagta gcgctcttcc 40620
gcttcctcgc tcaactgact cgtcgcgtcg gtcggtccgc tgcggcgagc ggatcatgct 40680
cactcaaaag cggaataacg gttatccaca gaatcagggg ataacgcagg aaagaacatg 40740
tgagcaaaa gcccagaaaa ggccaggaa cgtaaaaagg ccgcgttgtc ggcgttttcc 40800
cataggctcc gcccccctga cgagcatcac aaaaatcgac gctcaagta gaggtggcga 40860
aacccgagcg gactataaac ataccaggcg tttcccctcg gaagctccct cgtgcgctct 40920
cctgttccga cctgcgcgtg tacgggatac ctgtccgctc ttctcccttc gggaaagcgtg 40980
gcgcttttcc atagctcacg ctgtaggatc ctgactcgg ttaggttcgt tcgctccaag 41040
ctggggtgtg tgcacgaacc ccccggttcag ccgcagcgtc gcgccttacc cgttaactat 41100
cgctctgagt ccaaccgggt aagacacgac ttatcgccac tggcagcag cactggtaac 41160
aggattagca gagcgaggta tgtaggcggt gctacagagt tctgaagtg gtggcctaac 41220
tacggctaca ctagaaggac agtatttggt atctgcgctc tgcgtaagcc agttaccttc 41280
ggaaaaagag ttggtagctc ttgatccggc ccgctggtag cgtgggtttt 41340
tttgtttcga agcagcagat tacgcgcaga aaaaaaggat ctcaagaaga tcttttgatc 41400
tttctacgg ggtctgcagc tcaagtggaa gaaaactcac gttaaaggat ttgtgcatg 41460
agattatcaa aaaggatctt cacttagatc cttttaaat ttttaaatca 41520
cctaagaatg tatatgagta aacttggtct gacagttacc aatgcttaat cagttaggca 41580
acctatctag cgatctgctc atttcttca tccatagtgt ccgctccc cctcgtgtag 41640
ataactacga tacgggaggg cttaccatct ggcccagtg ctgcaatgat accgcgagac 41700
ccacgctcac cggtccagca tttatcagca ataaccagc cagccggaag ggcgcagcgc 41760
agaagtggtc ctgcaacttt atccgcctcc atccagtcta ttaattgttg ccgggaagct 41820
agagtgaagta gttcggcagt taatagtttg cgcaacgttg ttgccattgc tgcaggcatc 41880
gtggtgtcac gtcgctgct ttggtatggc tcattcagct ccggttccca acgtcaagg 41940
cgagttacat gatcccccat gttgtgcaaa aaagcgggta gtcctctcgg tctccgagc 42000
gttgctagaa gtaagtgtgc cgcagtgta tcaactcatg ttatggcagc actgcataag 42060
tctctactcg tcatgccatc cgtaaatgct ttttctgtga ctggtgagta ctcaaccaag 42120
tcattctgag aatagtgtat gcggcgaccg agttgctctt gccggcgctc aacacgggat 42180
aataccgcgc cacatagcac aactttaaaa gtgctcatca ttgaaaaaac ttcttcgggg 42240
cgaaaaactc caaggatctt accgctgttg agatccagtt cactcggtga cactcggtga 42300
cccaactgat ctccagcatc ttttacttcc accagcgttt ctgggtgagc aaaaacagga 42360
aggcaaatgt ccgcaaaaaa gggaataagg gcgacacgga aatgttgat actcatactc 42420
ttcttttttc aatattattg aagcatttat cagggtttat gtctctgaat cggatcacata 42480
tttgaatgta tttagaaaaa taacaaataa ggggttcgcc gcacatttcc ccgaaaaagt 42540
ccacatgcg ctcaagaaaa cattattatc atgacattaa cctataaaaa tagcggtatc 42600
acgagggcct ttgctcttca agaattcgcg gccgaatta accctcata aagggatccc 42660
tatagttagt cgtattatgc ggccgcgaat tctcatgttt gaccgccttat catcgat 42717

```

&lt;210&gt; 114

&lt;211&gt; 34071

&lt;212&gt; ADN

&lt;213&gt; Séquence artificielle



&lt;220&gt;

<223> Description de la séquence artificielle: insert  
d'ADN du cosmide a26G1 - brin codant

&lt;400&gt; 114

```

actgcagtcg ccggaatccg cgggtggaact acagcagccg ctggtgcgta tgggattgga 60
ctcgctcatg ggggtgcaat tacgcaaccc gatcgatacg gatctgcgcg tcttctgtgcc 120
catggtccga tttctagacg gccccagcgt tgcggaactg gccagggatc taagcgatct 180
aagcgccctc agcgaaacga cgaagggtggc gccggaacct gcggcgcgagg cctcggttcc 240
tgccctctcc taccctctca gcgcgggcca gcaggcgctt tgggtttatt accgaagcgc 300
gccggaaagt cccgcataca acatcgcggt gatcgcgcg gcgagaggcg ctttcgatcc 360
gcaggcggtg cgccgttcgc tgcaggacct ggtggatcgt catccggcgc tgcgaacgac 420
gattgcgagg agtgggcgcg cacccggtca aacggtccac agcagcgctc cgggtgattt 480
cgaagtgatc ccgtgttccg cggacgatga ggcggtgctg atcgacggcg tcttccacgc 540
gcccttcaat ctgcgggaaa actgtttccg ctgcgctctc ctgggtgcagt cgggggaagg 600
tcaggttctg gccatcgctg tgcatacatc cctcgccgac ttctggtcac tgctggtgat 660
gggtgatgaa ctcccgagta tctacctcgc gaggacagct ggcggtctcg ctgtcgcgcc 720
gccggtcgcg agcttcgcgc ctttctgtcg ctggcagaaac gaactgttgg ccggaaccga 780
ggcgagcgcg ctttggaaact actggtcctc gcagctttcc ggccagcttc cggttctgaa 840
tctccgctcg gatcgctcca gtccgcgggt gcagagtttc cggggaaact ctactcgtt 900
cgaatcgaaa cccgcgctga ctgcgaaact gaaggcgctc gcgcggcgcg agaaccgcac 960
gtgcatcgcg acgctgatgg cggcggttca agtctctc tcccggttga cctacaaga 1020
agagatcctg accggcaccc tcaccaacgg tcggaacgaa cgggaattcg ccgatctcgt 1080
cggatacttc gtgaatcccg taactcctgc aggagaactc tcaggcgatc cggatttcaa 1140
tacggtgctc gcccggtatc ggcaaacgct tctcgcgcg atcgagcacc aggaatcccc 1200
gtatgcccg atcgtggagc ggttgggtcc cggactcgcg gttctattcg tgctccagca 1260
gcctcatcgc attcccgaaat ccgtgcccgt catgttgggt cagtcggcg gtgcgatggc 1320
ctggggcagc ctcacactgg agtcctggc gatgccgctg cgacagagcc ggtttgacct 1380
ggatctgatg atggtcgaaa ccgatggagg cctctccgcc tttctgcaat acaacacgga 1440
catttttgat gctgccacga ttgaacgtct ctccctgcac ttcgcccgtc tgctggaagg 1500
aatcgcgagg aatcccgccct gtccagttgt cgatectacc ctgctgacaa cccgggaacc 1560
catccagctg ctcgaaagat ggaatgcgac cgcccgaggaa ttcccgtccc aatgcgtgca 1620
cgagctgttc gaagctcagg ttgagttgac gcccgacgac atcgcggtga gcttcgggtga 1680
gcagaatctg acatatcgcg aactcaacgg gagcgccaac cggatcgcg actatctccg 1740
ctcgcgcgcg gctggaccgg gcgaaatggt tggcatccat gtacgcgggt cgtctgaaa 1800
cgctcgaggg ctgttgggcy tctgaaggc cggcgcgcc tacgttccg tggaaaccga 1860
atatccggcg caacgtcttc ggtgatgct ggaagagacc aggcgggtcg ttgtgctgaa 1920
tgtcacggaa ctggaagtat ggacgcagcc cgacaccaat cgaacccgc tcgcgactcc 1980
cgccgatctc gcctatgtcc tgtacacctc cgggttcgac ggccggcgca aaggcggtga 2040
aatcacacac caggcgctcg tcaattttct ttcgtcgatg cggcatgag cgggcatcag 2100
cgacgcgcat acgctctcgc cctcacgac gttcatgttc gacatttccg cgtcgagat 2160
ctttttgccc ttgagcgccg gcgcgcgct cgtggtggcg aaccaggaga cggccgtga 2220
tggtgagagg ctggcgaggg aactcgcgcg cagcaaaagg acaatgatgc aggcaactcc 2280
cgccactcgg cgctctctgc tcgcatccgg ctggccggcg gaccgcgcgc tgacgycgct 2340
ctcgcgcggt gaagcccttc ctcgcatctc tgcgacccg ctccgtcaac gaaccgcggc 2400
gtcgaggaat ctttacggac ctaccgaaac gacaatttgg tccgcatcc aacgggtgac 2460
gacaggtgac ggaccggttt cgtatggcg ccccatcgca aacactcgct tctatgtcct 2520
tgacgatcgg atcgacccg caccatcgg tgttgcggcg gaactgtaca tcggcgcgcg 2580
cgggctcgcc cgtggatacc tgaatcgctc ggaactcagc gcggacaagt tctgcgcaa 2640

```

ttcgtttcgac cctcatggca ctccggtgta tcgcacggga gatctcgccc gccgccaaacg 2700  
 cgacggcgcg ctcgagatc tcggccggat cgaccaccag gtgaagatac gccgggttccg 2760  
 catcgaaacc ggcgagatcg agggccgggt ccgcagtcac ccggcggttcc gacatgctgt 2820  
 ggtcaccgcc agagaaaaat acgcggccgg taagtatctg gcggcctaca ttgtccccct 2880  
 tgcctgacggg catcgcgcgga cggcagccgc cgacacattc cagcagccgag tgcagttccga 2940  
 gcacgtgacg cagtggcaat ccgtctggga caccacatat gaacagaatg cgccgaacgc 3000  
 ggcacccggag ttcaacatcg tcggctggag aagcagtggt accggagagc cgattccagc 3060  
 tgccgagatg cgggagtggtg tgcaggattc cgtcgatcgc atctcgccct cgccggccgcg 3120  
 tcgctgtctc gagattggct gtggtacggg actgctgtct tcccgctgcg ctccccactg 3180  
 ttcggagtatc tgggccaccg acttttcgca gaaggcgctg gactacatcg ccgctcacgc 3240  
 ggaccgcacc ggctggcaa atgtccgcac gttccggcag gcggccgacg accgctgcga 3300  
 gatcgacagt cgctcgtgcg atgcggttgt tctgaactcc gttatccagt acttccccgg 3360  
 cgaagcgtat ctgcggcgcg tgcctggcga ggcggtgctg gtggtcaaac cgggcgccat 3420  
 cgtatttgtc ggcatgtcc gcagttctcc gctgctggag acgttttacg ctctcttaga 3480  
 agttcagcgc gcacccgct cgttgaccgg gaatgagttt cgggaacggg tgcgtttcgt 3540  
 cgcgtcgccg gaagaggaac tccgcgcttc ttctttgtct tcgcggcaaca 3600  
 gattccggag atcgcccgga tgaatactct gcccgctcgc ggccggctgcg ataacgagct 3660  
 gaccgcttc cgtaccagg gcatcctgca tatcgatcg cgggaagcgg aggagccgga 3720  
 atcggtatcg aggcgttgcc agaccggcgc cgaataacgc agagactcta gcgagctca 3780  
 gccggagttg gccgcattta ccgagattcc gaacgcacgg ttgaccgcgc aaagcccat 3840  
 tgtgacctgg atgaacggtg acgaagcttc agagacactc ggggagttgc gggaaccgct 3900  
 gcgcacagag tcgcttccg cgcctcgatcc cgcgatcta tggcgtatgg acgaagacct 3960  
 gccgtaccgc gtggcaatcg actggagcag tcatgggcca caggagcgtc tcgacgcgac 4020  
 ctctgcctg gtcggggcgc tccgcgcggc ttcccgctcg cgacgcgcgc tgcggcgccc 4080  
 gtatacgaac gatccgctgc gagccgtcta tacgcgcacg gttgtgcccg agtgcgctac 4140  
 tcatctgaag gagaagctgc ccgactacat gatcccgacc cgtggtgtcg tgcctccagc 4200  
 aatgccgctg acgcccacg gaaaaatcga ccgtaacgcc ctgcccgatc ccgagcccag 4260  
 ccggcgagcc cagccgaag cattcacgcc tccggaacct ccggtggaac aggtactcgc 4320  
 ccacatttgg ggcgaggtgc tcggcatgga tggcatcggc tcccatgac acttcttcga 4380  
 ctctggagga cattcgtcgc tggtaacgca gatgatcgcg ccgctgcgcg acatgctcca 4440  
 cgtggaagtg ccctttcgaa ccgtgtttaa cgccccacg gttcgaggct tgcgcgtcgc 4500  
 tattcaggac ggcgtagacc caggatgggc gcgagtttgc tgagctcgtg 4560  
 tttcccaatt tcagatgttc aaatcgagcg tatgatgagc gccgcccaag actaggaaca 4620  
 cagcgggttg taatcgaaat ttctgtcgca aataccatag acctctcgct ccccgccgcg 4680  
 caattgtctg accgtctgct gcaggaaaac agccccgaac atcgcatccc gcggcggtgaa 4740  
 aaccgggatg ccgcaccctt gtgcgtggcc cagcagcggc tttggtttct ccatcagctc 4800  
 gaccggattt ctcccgctta caacattccc atagcgtcg atatccgagg tccgctggat 4860  
 attcgcgtcc tctcgcgaag tctggaggcc gtggtgcagc ggacgcagag cctgcgcagc 4920  
 tgcattggcg gtgtggatg agaggcgccg caagacctcc tggcgcgagt gacactggaa 4980  
 ctccgggttg ttcaggctga cggaatcgca gaagcgccgc aaatggccct gcgtgatgcc 5040  
 cagatcccggt tgcacctcgc aaaaaccccg ctctcgcgga ccaagctgat ctgcctcgat 5100  
 gacaagcagc agattctctc gctgacgttg agccacatca tcgcggatgc gtggtcggtc 5160  
 gagacgttgc tccgcgacct gacgcgatcg tacgaagcgt tcgtgcaggg gcggccatcg 5220  
 ccgctcatgg aactccgat tcaagtatgg gactggggcg tccatcagca gacgtcgtg 5280  
 aaccaaaccg ccgacagcta ctggaagaaa cagctgtcgg gcaacttgcc tttctcgcat 5340  
 ctctctaccg atcgcgcccg gcccgccgag cagacctggc gggggcgccg ggagaccaca 5400  
 gccctcgccc gtgatttgac cgaatggact cagcggtttg ccttcgctga aggagcagc 5460  
 gtgtctatga ccgcaatcgc ggcgtttcag gtgctcgtac atcgtctata ccgcgcagaa 5520  
 gacatcctta tcgggggtcc agtcgcgggc cgtacacaa gagaaacgga aggtctcgtc 5580  
 ggtgttttcg ccaacatgat cgtcctgcgc ggcatctgc gcagcatcc gtcgtttcgc 5640

agtcttctcg cccgcacccg cgacaccgct ttgagcgccc tctctcatca ggactttcct 5700  
 ttcgaacgcc tgggttgagga actgcatacct ccgcgggacc tgagccgggtc gcttgtattt 5760  
 cagggtctcct tcgcgctgct gcccgatgcy ccggccatca ccgtcatgcc tgggctcacc 5820  
 atctcgcgcg agtacatgca caacggcgga tccaaactcg acctcggcgt gaccttcgag 5880  
 ccatccggcg atggactgat ggcgtccgcg gaatacaaca ccgatttgtt cgtatcgcca 5940  
 accatcgctc cctctgctga tcgctaccga acctcgtcgg cgagcgttgtt gacggatccc 6000  
 gacgtccgca tttcaacgcg tcgcctgttt tcccccgcgg tccgaagccg gatgctcgag 6060  
 cagcacaaat cgacacggcg cgaatgcgggt ccgaacgggt gtgcgcatga actggtcgaa 6120  
 gctcaggcgg aacgcactcc gcaacggctc gccgttgtct tcgaagacca tccgttgacc 6180  
 tacgcgcgag tgaatgcgcg ggccaacgcg ctggctcctc gtctgagcgc atccggcgcg 6240  
 ggcccgggaa agatcatcgc tctggcgatg gagcgtctcg tggagatggt gattgcgctg 6300  
 ctctcgattc tgaagtccgg cagcgcgtac ctgcctctcg atcccgcgca ccccaaggat 6360  
 cgtctcgcgc ggattctcga tgaagtgcga ccgcacgcgg tctctacgca ggagcggtg 6420  
 gctgagatga tggcgatgat ggcgatgatg gcggtccgcg tcgaaccgca agctcgcaat 6480  
 ctctgcgagc cgacgaaccc cgacgatctc gctacatca tatatacctc cggatcgacg 6540  
 gggcgaccga agggcggtgga gatccggcac tcgtcgctag tcaatctgct gcgctccatg 6600  
 cagcgcgagc cgggtctgac agccgcgcgt gggctggtcg ccgtcaccac cgtgtcatc 6660  
 gatattcgcg gactggagat ttggtcgcg ttgatcccg gcgcgcgctc catcgtcgcc 6720  
 acccgcgaga tcgtggttga cggcgagcgg ctaccaccc tgcgtgataa gtcggcgct 6780  
 accgtcatcg agcgacccc gagcgggttg cggcaattgc tggattcggg tgggaagccg 6840  
 ggtaaaggct tcgctgtttt ctgcggcggt gaagctctgc ccgcggaact ggcgcgcgcg 6900  
 attctcgata gtggcgtaga gctctggaat cttaccgga cgacgcgagac caccatattg 6960  
 tcggcgctgc acaagacaca aagactgggt gcctccgata gcatcgtgcc gatcggccat 7020  
 cccatcgaca acacgcagtt atacatcctg gattcgcgca tggagccggt tcccccgga 7080  
 gttccgggag agctgtacat cggaggagcg ggactggcg gggcgctatca tcgcaacccc 7140  
 gagctcagcg gtgagaaatt ccgcgagtgag cgtgatcgag gacgcattta ctacacgcg 7200  
 gatctggctc gctaccgttc cgacggcgca gtcgagtgcc tgggacgagt cgatcgccag 7260  
 atcaagctgc gcgggttcg catcgaaccg gccgagattg agggccgcat cgagacgcac 7320  
 attgcccgtga agcaggcgat tacggtcgtg aaggacgac gtctgatcgc ctatctcgtt 7380  
 ccggcaacgg gcgacgtgcg cgaatctgag agcgatttgc ggtcgtggct ggcaacgcgcg 7440  
 ctccccgatt acatgatccc ctgcggcttt gtcagcctgt cctcccttcc gtcgacgcc 7500  
 aacggcaaaa tcgacgcgaa cgcgcctccc ggtttgccca caacgcgcgt tctgtctcgc 7560  
 gagccgatgc gcggcgatgc ggtggagacg attgcgtcca tctggcgta agttctcgcg 7620  
 gtgagcacag tcgactatgc cgagaaactc ttgatgtcg gcggcgactc gctaatgctc 7680  
 acacgggtgc gcggactgct cgaggagcgc ctggggttga cgtctccgt cgtcgatctg 7740  
 ttccggcata cgaatcgca gtcgcttgcc ggcctggcag aaaaatccga acccgccgct 7800  
 gcggaacact gcgctgcggg ccgagaagat ttatcgcaag ttatcgcat ggccggccgg 7860  
 ttcccggggg gcgcgaatg ggaggagttc tggcgcaatc tgcgcgacgg tgtggaattc 7920  
 atcgccaagg tttcccgga agatctcgtg gcggcgcca tcagcccgga ggtctccag 7980  
 gacccgagct acgtgcgggc caaagggtctg ctggagcgca tcgagttttt cgatgcgcg 8040  
 ttcttcggct acagtcgcgg cgaagcggag atcatggacc ccgacatcg cgtgtttctc 8100  
 gagtgcgcgt gggaaacgat ggagaaacgc ggatatcgcg ccggaagctc taagggtcg 8160  
 atcgcgcttt tcgcgggatg cggcgtcaat acctactcgc tgaacaaact cgccaccgcg 8220  
 gagcggcttg atttctcag cccctccgcg taaccagctg tgaaggccaa cgacaaggat 8280  
 ttcttggcca cgcgtgtctc ttacaagctg aaacctccgc gggccagcct gacggttcag 8340  
 accggtgctc ccactcgcg ggtgtcgggt gtgatggcat gcgagagctt gacgcgcggc 8400  
 gcctcggaca ttgccttggc cgggggagtt gccatcaat tccgcagtc cgtggggtac 8460  
 ctgcaccagc cgggcatgat cctgtgcgcc gacgggcgtc gcgcgcgtcc cgatgattc 8520  
 gctcaaggca cgggtccggg caacggcgcg ggtgtggtcg tctcaagcg cttgagcgcg 8580  
 gctctggcgc atggcgacac gatctacgcc gtcattcgcg gagcggtat taataatga 8640

ggcgcgcgagc gcatgggggtt taccgctcca ggtgtggacg gtcagacgcg attgattcgg 8700  
 cgcactcaag agatggcgcg cgtgaagccg gattccatcg gctacatcga ggcccaacga 8760  
 acagccacgc cgcctcggcga tccggtggag atcgccgccca tccgtcgcaa ctctccgaaa 8820  
 aacggaaagcg gcgatgtgta tatcggaatcc gtcaagacca acatcgggtca tctagacgtc 8880  
 gcggccgggtg tggccgggct gatcaagacg gtgcttgccc tccatcgccg ccgatttccct 8940  
 cccagcctga atttccagcg tccgaatccg cgaattgatt tccgcaaacac tccggttccgt 9000  
 gtgagtaacg ggctgctcga ctggcccgcc gaaagaccc cgagacgagc ggcagtcagt 9060  
 tcggttcggga tggcgccgac caacgctcac gtgattctcg agcaagcgcc gcccggtgacg 9120  
 ccggcccgag ctggcccgca acgatccgca catgtgcttt gcctgtccgc caatacacag 9180  
 gcggccctcg aagaactggc gcgctcgat cgcggccata tggacaacca gcccggtttg 9240  
 tcggttcggcg atgtcgcatt cacggccaat gcaggggcg tgcacttccc gcaccgtatc 9300  
 tgcattgtgg ccgggtcgag cgacgaggct gcgcaacgac tgacggaggc acgacggggt 9360  
 cgcacgcgcc agacgcgccc caagattgcg tttcttttca ccgggcaagg tgcgcaatac 9420  
 gcgggcatgg gccgccagtt ctacgagtcg cagccgggtt ttcgcccgcg catggatgaa 9480  
 tgcgcagtcg tgcggaatgg acggctcgat ctgcccggcg tcttgccga tgacgcggtt 9540  
 ctgcagcga cgcggcgcc gcagcccgcg ctggttgcct tgcagtgggc ctgggcgca 9600  
 tctgtggaagt cctgggggtg gacgcccgac ctggtgatgg gacacagcgt cggcgaaatac 9660  
 gcggcccggt gtattcgccg cgcgctcagc ctgcccggat gcctcggtct agttgccga 9720  
 cgcggccggc tcatgcaaaa cctgccggaa ggtgcgatgg ctgcccga cgcggcgag 9780  
 cagcgctgtg ccgcagcgat caactcgccg gcttccattg cggccatcaa gcgacccgct 9840  
 gaggtcgtag ttttcgggtg gccgacggat attgagagcg cgcctggcaac tctacgtcgc 9900  
 gagggatcaa aaacgcagat gctggccggt gcgcgcgct ttcacagctc gagcatggat 9960  
 ccgattctgg cggactcgca acgcccggcg gcggcgatcg cgtggcgcaa tccctcgatc 10020  
 ggcctggttt cgaacctcac gggcaaaact gccggcgagg gacagctggc gaatccgctg 10080  
 tactggcgag atcacgctg aaacctgtgc cgtttcgccg acggtatcca aacgcctcaag 10140  
 gacgaaggct gcagactgtt tctcgagatc ggttcttacc cggcttaccg cggcatgggc 10200  
 caaaagtgcc tgcgcgacga cgccaagcag tggctcgctc cgtcgcttaa agggccgcat 10260  
 gagtgggaga cgatcttcag cagtgtggcg acgctatatc aggggtgggtt cgacatcgat 10320  
 tggcaggagt tgcaccgtcc gtattcgca aggcgtgtcg cctgcggcg cctcatcttc 10380  
 gagagacgcc gccattggat cgagcggagt tccagaccgg aacctgtagc ggttcgcatg 10440  
 ggtctcgctg ggtgccgggt gtcgctaccg ttggcgagac ttatcttcga gtccgaaacta 10500  
 tgcagcgctt cgcctctact ctacagacc ccatattacg gttecggtgtt gggcccgccc 10560  
 gtgtacttcc tggccatggc gctcgaggcg tccggcgagg tgtttggcgc gcgcccgcac 10620  
 accggtgaaa acgtgaactt cgcgcacct ctgatccttt cagtccgctc cgacacgcct 10680  
 gtccagctgt tgccttcaca gagcgatgac cggcatgcct cgttccgcat actcagcttg 10740  
 tccagacgct cgtggaaact acatgctgcc ggaatattg ccgccacgc tgggtgtcgt 10800  
 ccggttcccc gactggtcga tgaacccggc cctcggtgg atggagacac gctattatcg 10860  
 cctgtcgccc acctcgagat agaactgggg ccgagctacc gccgataca cgcctatcat 10920  
 ttcggtgaac aggaagcgct ggcccgatg gattccgcaa cgcctctcaa tcccggtgt 10980  
 gaattggcgg aagccggcct gcaattgctt agcccgccg cgagctccgc gcttgcggat 11040  
 ggcccgcaac atccgatatt cgcctccgctc ggtatcgatc gcgtttggtt ttacggcagc 11100  
 ctggaggggc ccgatgtggg gcccgcgcaa attctccggc attccgccc ggcgtttacc 11160  
 ggccaggcgc agttctcgga ctccggaggc tgcgttctcg gggaaactca gggcgtagt 11220  
 ttcggcgcg ctactcgcc atgggcgcag cctcggaac ggaagcccca attgtatgag 11280  
 gtcgagtgcc ggcccgcaac gctccgcca cgtcacagcc tggggcatgg 11340  
 ctgatcctgg ccgacagtgg cgcgcggccc cgcgctctg catagtcgct cacagctcag 11400  
 ggcgagtagt gcgttacctg gccgcagcc ggcgagtag tgcctcagc cgtgtagcgt 11460  
 gactggcgcg ggaatcgtaa cctgtacagt ctcgatgatt atgagctcgg ctgcgcgacg 11520  
 actcggccce tggtagaagtc cctgaagtcc ggtccggcg tatggctggt aacggccgcg 11580  
 gcgcaggcga ccagtcgggt gcacaatccc atgcaggcg cgtctgggg ctctggcggg 11640

gtgatcgcg cgagcacc ccggtctgtg ggcgggctca tcgatctgga tcccgacgat 11700  
 ggcgatgctt cggcgccg cgccgcccgc cagatgcgtg atttcgacgg cgaagatcag 11760  
 tcggcgctgga gaagcaacc gcgctacgtg ccgcgactga cccgcgcgacc cagcgcgcca 11820  
 cgccgacgtc gtctgggttc gggcgcgact tatttgatca ccggcgggctt cggagccctg 11880  
 ggaactacag tcgcgaaatg gatgggtgag cacggcgcca ctgcgctcgt cctggccggg 11940  
 cgccggccctc caaacgagga gcagcagcgc gtgctgcaac agatttggtgc gacggcgagag 12000  
 acggtcgagc tcagccggga agaagaggtc ggggatctca ttgcgcgcgt ccacacccgaa 12060  
 acgtcaccgc tgcgcggcgt tatccatgcc ggggggtgtg tcggacgacgg cgtactcgtg 12120  
 aatcaggact ggacgcggat cgcgaagctc atggcgccga aggcggaagg cgtctgtacc 12180  
 ctccatcatc acaccgcgga tctgcgcgtc gacttcttgc tgctcttttc atcgccatcc 12240  
 tcgctcttag gtccctgccc gcaggcgagg tacgcccggg ccaacgcgctt tctcgatgag 12300  
 ctggcgcatc accggcgccg actgggtttg ccggcgacca gcattaaactg gggggcgctg 12360  
 tcgggagccg gaatggccgc gcgcaccagc cagtctgatg ccggcgctggc gagcctctcc 12420  
 gtggacgagc gctctacatc tctcgaggcc gtccctgatg aatgcccatc tcagattgcc 12480  
 cgcctaccgg cgggctcgat taccggcgag ttgctgcgtc ccgcgcgctt gccttcactc 12540  
 caactgogca cccgcttgaa cgaagccaca ccccggcgag cgaagaccat cctcatttgc 12600  
 cacatcaggg agtcaactgg cgctcttctc ggcctcgga cttccacacc gctcgatcca 12660  
 cagcagctct tgggtgaaat gggactcgat tcgtaaatg ccatagacgt tcgcaactcg 12720  
 ctctcccaat cactggggca gcctttgccc gogagtctgc tggctgacta tccgtcgtc 12780  
 gatgogctgc tcagttacgt gctccatgag gtatttccac ccgaagcatc accggtggaa 12840  
 gcgcggaggt tctgagaact cgcccccga gaaactggaa gactgctcga ttcgcggctg 12900  
 gcgcaggtcg accagtggtt ggagacgcaa taaactgag cgggtcagac gatctcagca 12960  
 agctcgcgcg cgccgtgatt gcgctcgaca aggtgcagaa aggtctgac cagctggaga 13020  
 gcgcgcgcag cagaccatc gccctcatgc gcgcgggctg ccgcttcccc ggcgcatcca 13080  
 atctcgatgc ctatttggtg ttgctgcgcg agggcgccgag ccgctgacgt gaagttccac 13140  
 ccgacgcgct ggacatcgat gcctactacg atccggatcc cggcgccgac ggccaatgt 13200  
 acacgcgcta cggcgcttc atcgatcagg ttgaccgttt tgacgcccgg ttcttcggca 13260  
 tcgctccgcg cgaggcgatc agcctggatc cacagcagcg gctgcttctg gaagtcaact 13320  
 gggaggcgat cgagaacgcc gggcttccac ccgaccggct ggcggggagc cggaccggcg 13380  
 tcttcatggg gatcttttcc aacgattatt acaacctgca aatgcgcggc ggggatgcgc 13440  
 atatcgagc gtacacccgc accggccaata cggccagcgt tgccgcggg gctctctgt 13500  
 acatcctcgg gctgcagggc ccgaacatgg ccatcgacac ggcactgctg tcatcgctg 13560  
 tcgcgggtga ccttgctctg cagagcctgc gctcaggtga aagcgacctc gcgctggcg 13620  
 gcggcgctca tctgattctc tcgcggatc ttctgcgact ctctgcgaag ctgaaggcga 13680  
 tggcagccga cggctcgtgt aaggcattcg atgcgcgag agacggctac gtcgcgggtg 13740  
 agggctcggt tgtgtgtgt ctgaagcgac tctccgacgc gctgcgcgat cgcgatccgg 13800  
 tgaatggcgt gatctcgccc accgcaatca accagcagc acgcagcaat ggaactgacgg 13860  
 gcgcgaacgg gccgcgacag gaagccggtga tccgccaggc tgtggggagc gcgcgcttgc 13920  
 agacgctgga tgtgactat gtcgaggcgc accggaacgg caccgcgctg ggogattcca 13980  
 tcgaagccgg agcccttgcg gcgcgcgtgg gagcggggcg caccacggcg aaccaagctga 14040  
 agctcgggtc ggtgaagacc aacttcggcc acctcgaggc ggcacggcgt gtggcgccac 14100  
 tgaatcaagt ggcgctgctc ctgcagaacg aagccattcc gcccatctgc aactcgacca 14160  
 gcgccagccc gcacatcgat tggaaacacgc ttccctcga aatcccgcca cggctcacc 14220  
 cctggcgggt tgcaccgcgc gggcgccgag tcgccggcat caactcgttc ggcctgagcg 14280  
 gtacgaatgc cgcagctgctc atcgagcagg ccgcgcaaca ggcgcgctcc agtaccggcg 14340  
 caccgtacct gcttcgcgta tcggcgcgca gtcgcgaggc gtcgctgat ctggcgcgcg 14400  
 catacgcgca cgtggtgaac gacaacccc gcacacctg ctacacggcg tgcgctcgcc 14460  
 gcacttcata cgaacacgcg gcggcattca ccgggacgaa ccgcgaggac ttgagggcg 14520  
 ggcggacag ttttctggcg ggcaacccga accgcgatac ccgcaacaggt tttgtgcgcg 14580  
 gcggcgagaa gcgaaaagtc gttttctgtt tgccgggaca aggatcgagc tggcccgcca 14640

tgggcccgcga cctgatggct tctgaaccgg tgttccgtgc cggcatcgaa gaggtcggcc 14700  
 ggcgcatgca gccctacgtc gactggctgc tgacgcaaga gttgcagggg ccgctcgacc 14760  
 gcacgcagct gattcaaccg gccctgttcc cagtcggggg cgccttggcc ggaactgtgc 14820  
 gccattgggg aatcgagccg gacgcgctga tcggccacag catgggcgaa gtgcgcgacg 14880  
 cgcacattgc aggtgcgctg actctcgatc aagccgctcg ggtgatttgc ctgcgcagcc 14940  
 gggatctcgc cggagtagcg gggcaggggag aaatggctgt cgtggaatta gcgctggagc 15000  
 aggcctcgcg tggcatcgcc gggcgctcgg atcgggtctc gattgcccgc agcaacagcc 15060  
 cgcgcagcac cgtcctgtcg ggcgacagcg cagctctggg cgaactgtct cgggaaactgg 15120  
 aggcgaagaa cgtcttctgc cgtcgctga aagtggaat tgctcgcac agccatctga 15180  
 tggactccgt tgcgcggcg ttcgcggcg cgttcagccg cggccggccg 15240  
 cccttggcat gtactccacc gtcaccggcg cagcgattag cggtagaag ctggtttctg 15300  
 cgtactgggc tcgtaattct cgccaaaccg tgatgtctgc gacggcctgc gccgcagccg 15360  
 cggcggtggg tcatgatgtg ttcttggaac tgagtcacca cccgttggg gtccagccga 15420  
 tccaggaaac gctcggagat cgggcagcga ttcgcgctgc ctctgttcgg cgcagtgaag 15480  
 acggaaacct gcactcgcg cggcagctgc gagcgctgct gactaacgga gtcactccgg 15540  
 actgggtctcg tatttatccc aacggcgccc aaactcgcg gctgcaccaac tatccctggc 15600  
 agcgtgagcg tcatgtgatc gatatcgct cgcccgaggt cagatctcag gcttgcctg 15660  
 gcggcgagat cccgtgcgcg ctgcgggaga tgcagttcga gtcactgtg gaggcgaaag 15720  
 atttcgcgga tcaccggctg cagcatgtga tcgtactcc gggagcgtgg cacctggcaa 15780  
 tggcgctcgc cgtcgcgcgc caaggtctcg gcgcggggcc tcaccatgtc gaacacgtgt 15840  
 catcgagcgg cgcgctgacg ctgcgggaaa acgatgctgc caggcaggtt caactgggtac 15900  
 tcgctcatga agaaggcgcg ggagcttctc tcgcacatca cagccgcgag gattctctga 15960  
 agctgcacag cgaaggcatg ctgcaggcgg gcgattccac ggcattccat gatctggatg 16020  
 cgattcgcgc ccgctgcacg gcggagctca cagccgatgc cttctatttc cactgtatgg 16080  
 atcgcggcta tcaactcggg cccaccttcc gaacctcgg ccccatctgg ccgcggcaacg 16140  
 gtgaggtgct ttgtcgcgtg gacattccgc tgacggaaat cgagacgatc gactgctgtc 16200  
 tgcagttgcc cgcggccctc gtcctcacg acgattgaa agatgtgcat gtcccggtag 16260  
 gtctggagcc attctcgtc gctgaagtgc ccaactggccc ggtctgggga tacgcggtct 16320  
 tgcggccgga ttccacgggt gatgtcgtc tcgtcaccg caccggcagc gtggtggcg 16380  
 aatctgttgg gctgcagtc agagtgcgcc atagcgcca cgtccgcgaa tcggagattc 16440  
 ccactcggac ggtgcaatgg acccgctcgg ttcgcggcg cgaagccaat gccggcaatg 16500  
 ctggcggacc ttggctcgtc atcggcgagc cggcgattgc cgagactctg caaaaagcgc 16560  
 gccaaacctc ccgcacggcc gatagctgtc cgggtccgc gtgccgtcaa attgtgact 16620  
 gtccctcgcg gcgcacgcac gacctcgtt ccgtatttgc cagcactcgt caagcgggct 16680  
 ggctcgagcc gccgcgcctg tggctgctga cgcgcggatc tgccgcggtt ctcaactccg 16740  
 acaaaagata tgatatcca caagctggc tgcaaggaa tgggcggagc attgcctatg 16800  
 agcatccgga gctgcgctgc acgctcgtc atctcgatgc gcacagcaac gactcgggcg 16860  
 atctcgcgac gctgatgctg tcgaatatcg cagaggatca agttcgcatc cggcaaggca 16920  
 cggtagggcg ccgcgcctc agtcttccca agatcccat gcaccgatg gtggcttcc 16980  
 gtgcgcagc aaacctatct atcacggcg ggcctcggcg actcggactg caggtggcg 17040  
 gatggctcgc ccgcgcggga gcgcgcacat tcgttctgct gggacgcagc gaggctctc 17100  
 ggcccaactc ggaagtgctc aacgtcaaga tcatctatgc ggaactcgc cgcggcagc 17160  
 agctatcggga tgcgctcgcg atcatcgatc gcgacatgcc ccgcttgcgg ggcgtgttcc 17220  
 atctggcagg caacctggcc gacggcatgc tgcctaatct cacgaccgaa cgtctcgaag 17280  
 gcccatggc tcgaaagata gcggcgcggt ggaacctcgc cgaactcacc ccgcgcggcg 17340  
 cgctggatca ttttgtctc ttctcttccg ccagcgcgac agtgggattc ccgcgcagc 17400  
 gcaactcgc ccgcggcaat tcaattctc acgcgctggc tcactctgc cgcgcceagg 17460  
 gtcttccgcg cgtcagcatc gcgtggggac cgtggacaca ggttggttgt gccgcacagg 17520  
 cgaaccgcgg agacctcgtg gccgcgcgag gcatctcggt tattcaaccg caacagggat 17580  
 tgcgcgcgct atcaaaagca ttgaacgaga ttcggccgca cgtcgtctc atgaacttcg 17640

|             |             |             |             |             |             |       |
|-------------|-------------|-------------|-------------|-------------|-------------|-------|
| atategcgca  | gtggctccgt  | tactatccgt  | cggcgcgcatc | gatgtccctg  | ctggccggca  | 17700 |
| tcgcaccgc   | ggccgcggac  | accaaaccgg  | cggccgcacat | gcgcagcgag  | ctcctggcag  | 17760 |
| ttccagccgg  | gcggcagcgc  | cgcgcgcggc  | tggaaacgct  | gctgatgcac  | gaagccggac  | 17820 |
| acgtgctggt  | cttcgatcca  | gcgaacatcg  | acggcagagc  | gacgctgggt  | gatctcgagt  | 17880 |
| tcgattcggt  | gatggccctc  | gagttctcgca | acgctctgga  | agccgggctg  | cgcgtcaagc  | 17940 |
| ttctcgccac  | ctgatattcg  | cgttaccoga  | cattctccgc  | cctggcgagc  | catctcgccg  | 18000 |
| acaagctcgg  | cctgcgcgtg  | gaaagcatgg  | cgggcaatgc  | tgaaccttcg  | accgttgctg  | 18060 |
| ccgttgctac  | ccttgctacc  | gttggcaccc  | cgcggggcga  | ggaccggagt  | cccgcgcgtg  | 18120 |
| cagacgatct  | cgacgcgctc  | gcaaaccaga  | tcgcgggggt  | gggggacaaa  | gaaatcgaag  | 18180 |
| cttttgtaga  | acagaagttc  | gctcattttt  | caggagcctc  | cgagttagtt  | cgatatccga  | 18240 |
| gcgattcccc  | aaccttacgc  | cgttgacgca  | ggcgtaacct  | acgtcggagc  | acatgcagcg  | 18300 |
| acgtctcgat  | gcggccgaac  | gcgacgcgcg  | cgaacccatc  | gcgatcgtgg  | gtctggggctg | 18360 |
| ccggtttccg  | ggcggcgatg  | ggcccgatga  | gttctggcag  | atgttgccga  | gtggagtcga  | 18420 |
| tgctattcgt  | gaggtaccgc  | ctggacgatg  | ggacgaaggag | tcggtccggc  | gcactcctgaa | 18480 |
| atcgcttccc  | cccgcacgcg  | cggtagaatg  | tcaagccgga  | tttctcgatt  | ccatcgtatgg | 18540 |
| tttcgacaac  | gatttttttcg | gcattttcgc  | acgcgaggcc  | gtcagcattg  | atcccgagca  | 18600 |
| gcggctcgtg  | tgtgaagtgg  | cgtgggaggc  | actggaggat  | gcggggcaga  | cgatgggaag  | 18660 |
| gcctctccgc  | agccgcacgg  | gcgtcttcgt  | cgggatccac  | agccaaagca  | cgactatttt  | 18720 |
| ctggatcgag  | accgcccgatg | gcgcgcgcac  | cgatccgcat  | accgccaccg  | gcacggcgca  | 18780 |
| tagcgtgate  | gcgcggcgac  | tttctatttt  | gctgaacttg  | caaggaccac  | gcactcgctg  | 18840 |
| cgacacggcc  | tgtctgtctt  | cgctggcgcc  | gggtcatctg  | gcgtgcccga  | gcctgcgcag  | 18900 |
| cggcgagtag  | acgctggccc  | tggccggcgg  | agtgaattctg | cgtctctcgc  | cggagtattat | 18960 |
| gtacgccaac  | tgaagatagg  | gaaaccgcct  | gccacgcggt  | cgcctccgcg  | cttcgaagc   | 19020 |
| ggcggcgagc  | ggcatcgtgt  | tcggagaagg  | gtgcggcgtg  | gtgggtctga  | agccctcgtc  | 19080 |
| cgatgcactc  | gcggccggag  | accgggtgtg  | ggccgtgggt  | cgcgctcccg  | cgctcaatca  | 19140 |
| ggatggccgc  | tcggccgggc  | tcacgcctcc  | caatgtcgtg  | tctcagcagc  | tcgtcatccg  | 19200 |
| gtcggcattg  | gcacatcgcg  | gcgtccgcgc  | gcagcagatc  | gggtacatcg  | aagcccatgg  | 19260 |
| caeggggaact | ccgctcggcg  | atcccatcga  | gatcagaggc  | ctggcggaaa  | ccgtcgccct  | 19320 |
| cccgcgacct  | gtcggcgatg  | tgtgcgcggt  | cgggtccctg  | aaatcgaaac  | tcggccacct  | 19380 |
| ggaggggagc  | gcaggcatag  | cgggattgat  | taaagcgggt  | ctcgcattga  | gtcacgagac  | 19440 |
| gataccgcgc  | agctttacacg | tgagacagct  | gaaccggaat  | atccggttgg  | agggaaactc  | 19500 |
| gctcgacatt  | gtgaaggaa   | tcggcccggt  | gcccgcgggt  | tcgagacgaa  | ggttttcggg  | 19560 |
| cgtcagcgcg  | tttggttggt  | coggcacgaa  | cgcgatctgc  | gttcttgaag  | aaagccggcc  | 19620 |
| gactgctaga  | ggcgaagctg  | cgagcggggt  | ccattcccca  | ccccccgcgc  | ccgctgcgcg  | 19680 |
| ggcggctgtc  | ccccctcgcg  | agggggacac  | tgggggcact  | cccgaacatt  | caggcaactc  | 19740 |
| cgacactgca  | gacactcccg  | acactgcaga  | catctccgac  | atttcaggga  | ctcaggcacc  | 19800 |
| tgcggcaact  | acgggcatctg | cagacgcgat  | gtatgtgctt  | ccgctgtccg  | cgcagtgtgc  | 19860 |
| ggacgaactg  | cgtcgggtgg  | cgcgggcata  | cggggaattg  | ctgcacagct  | cgcacgcacc  | 19920 |
| gagctcgctg  | gatctttgct  | acacggccgc  | aggtcccgct  | acgcataccc  | gatcggcggt  | 19980 |
| cgtctgttcc  | ggcagaacgg  | ctgaagaact  | ggcggcgcg   | ctccagggga  | tcacgatccc  | 20040 |
| ttccacgagt  | cggaagacgg  | tattcgtctt  | ctcgggacag  | ggatgcgaat  | ggatcggaat  | 20100 |
| ggggcgcgag  | tggatggacc  | gcgaaccgct  | tatttcgcag  | gcgttggaa   | cgcgtagggc  | 20160 |
| cgcctatcgg  | ccttatgtgg  | actggtcgtc  | gaaagaagaa  | ctggcggaagc | tcgaccgcgt  | 20220 |
| cgaggtcaat  | cagcctgcgc  | tcttcgcgct  | gcaggtcgcc  | atcgccgcac  | tgtggcgctc  | 20280 |
| ctggggaaat  | gagccggatg  | cgtctcagcg  | gcacagcatg  | ggagaggtcg  | ccgcgctcga  | 20340 |
| tgtcgcgggt  | gcgctgagcg  | tgcaggatgc  | ggcgcgggatc | atttcgagcc  | gcacggcgct  | 20400 |
| gttgagccgg  | atcagcgcc   | tggggcggaat | ggcgatgggt  | gagctgcgcg  | tcgcggaatg  | 20460 |
| tgaggtccgtg | ctgtcgactt  | acacggaaag  | actatcgccc  | gcggttctga  | acgggaccaa  | 20520 |
| ctccaccgtc  | atctcgggtg  | aagtcgaagc  | cctggccgag  | gtcgtcgcca  | cgtcggagcg  | 20580 |
| gcgaggcgtg  | tcttcggccg  | cggtagaaat  | ggacttcgcc  | gcgcatagcc  | cgcgaagtga  | 20640 |

cccattgtgc gacgaactcc tgcagtcgct cgacgggatt caaccgcggc ccgcgaccat 20700  
 acctttttac tccacgggtga ccggcgcgac gctggagacc accagctcgc acagcacgta 20760  
 ctgggctcgc aatctcgcat ccgcgggttct gttctggcag ggcattccgc atcttgccga 20820  
 cagcggggcac gatgtcttcc toagatcacg cccctcatccc atcctgctgc ccgcctatcgg 20880  
 cggcaatcgc gcgctgggttc cgtctctgcg ccgcgaccag gacgaacgcg gttccatgct 20940  
 cactgcgctg ggcgcctctc atgaggctgg gcacactgtc gcatggcgga ccgtgtaccc 21000  
 ttccggcaat tgcgtgcgcc tgccccggta tccctggcag cgtcgtcgtt tctggctcga 21060  
 cgcttcccc gcgcgacacg cgatecacgtt gggcaatccg ctggttggaa aacgcgtcga 21120  
 agcctcgacg caaccgggca ctttcttctg ggagacggaa ctcagctcgc cttccgtgcc 21180  
 ttggctggca gaccatcgcg tgcagggcga agtctcttgc ccggctactg cgtatctcga 21240  
 tatgctctcg gccggaaact ccgagacactt cgggtgaaagt ccgtgcgtgc tggagcatgt 21300  
 gactttcaca cagatgtcga ttgtgccgcg cgacgggcagc atgacgttgc agctggccat 21360  
 cgcgctcgat agaccgggga tggcgctcgt tcggatttcc agccggcgagg catcgacatg 21420  
 ggtctctgat gcttccgggg acattctgtca gacgctcgc gatgcacga ccgtcccgcc 21480  
 cgcgctcgag gacagcgtgc agggcccgctg cccacacagt gtgcggcgcg ccgagctgtg 21540  
 gcgtcagatg gcggagcacg gcgtcgagta tgggtccggct ttcgcgcgag ccagacatga 21600  
 ctggagttgt ccaggtagagg cgatecggcg cgtcgtatgc tcggaaacgc gttccactgc 21660  
 gccgcgcttc cgtcgatcag gtctgcagat catcgccgcg gcgtttggtc ccgcggctgg 21720  
 aacctggctg ccgcgcggca tcgaccggat gcgctggctg catcccgca gttccgtggt 21780  
 gtggagcgat gcgcggcttg aaggacctat ccgcgatctg tcgctgctgg acggagaggg 21840  
 acaactgggt ccgccgatcg agggctctgc gctgcagcgc gctgagtcgt ccgagcgcat 21900  
 cgacatcgcg ggtcgtgtgc acgaactcgc ctgggtcgct cagccgcagc ccgtcgcaga 21960  
 gccgcggcg gcgcgagcgc gcgggtcatg gctcatttgc ggcctgtggt atagcgcgt 22020  
 caccgcgatg ctgcgcgcta ccggcaaccg cgtgacgcag acctcgcggg aaaagctcga 22080  
 tgaactccag ccgccgctcg aggaatcgt gtttttgcgc gacgacgaac cctcatgcga 22140  
 ccgcatctcg catctctccc agaccctggg gcgcagcccc tggcgctcaag caccgcgct 22200  
 atggctggct acgcgcggcg cgacgccgt cgatggacag atctcgcaag ccggtatcgc 22260  
 tcaggcgctt ttctgggggt tgggcgggac cgtgcattac gaacatccgg aactgaactg 22320  
 cacgctgatc gatctcgatc ccgcgcggcg cgaagaggaa ctctcgcaag aactgctgac 22380  
 gaacaacggc gaagaataaa tcgcctttcg cggcgcgcg cgttacctgc ccgcgctggc 22440  
 tcggcacgaa gcggatatgc aaccgcccat gttcaaggcc ggcatcgcc cgttccggct 22500  
 cgagatcgat gcccccgag tcctcgacgg gctgcgcttgc cggggccacat ccgcgcgcgc 22560  
 ccgcgaagcc ggtgaagtgg agattgaagt ctgcgcgcgc ggcctgaact tcctcgacgt 22620  
 tctcgtcgcc ctcgcggtta tgcccgaaga tgcgcccgcc gcgtttgcgc gacgcgcgcg 22680  
 cctggcgccg gaatgctcgg gccgtatcgt ggccatgggg aaggcgctca ccgacttctg 22740  
 catcggagat gaagtctgtg cccctgcgcc ttgcagtttc ggtcgcttgc tcaccacgcc 22800  
 cgctctccgc gtgtccttga agccgcgcaa catctccgcc caacaggccc ccgcctcgc 22860  
 tatcgcgttt ctcaccgcgc attacgcgct ctgcgcgagc gcgcggcttg ccgcccgcgga 22920  
 acgagctcgt attacgctg ccaccggcggt tgtgggattg gcggcaatcc agatcgaca 22980  
 gcgtgcgggc gcggagatct tcgctactgc cgggagtcgg gaaaaacgag cgtatctcgc 23040  
 ctcgtggggc atcgcgcatc ttctggattc gcgtctgatg gctttctgct acgacatccg 23100  
 caattggagc aatcaagaag gagtgaagct cgtcctgaat tgcctttccg gcgactcgt 23160  
 ggaggcgagc ttgatctcgc tgccgatca tggacgggtc atcgagatcg gcaagcgga 23220  
 ttactatgcc ggcgcgaagc tggggcttgc cccgttctgt aagaacctct cgtacacgct 23280  
 ggtcagtttg ctcgcgatgt cctgaagcg accggcatgt accggcgatg tctgcaggga 23340  
 gatggtcgca aatctcgaat cggaaacctg gcggccctgc gaaacgcgag tgacgacct 23400  
 caccgaatcg gtagggcggt ttccgacat gcgcgagcg cggcacatcg gcaaaatcgt 23460  
 catggcgatg cgagattcgc ccaatgcgcc catcgaccc ctagctcgg cgttcgatag 23520  
 cgagggaaac tacttgatta ccggcggaact tggcgggctc ggtcttacgc tcgcacgctg 23580  
 gatgacgga cgcgcgcccc gggcgctggt gctgctgagc cgccgcgcgc cttcaccga 23640



ggtccagcaa gccatcgccg tcatggacgc agatgtccgg acggtgcagg ccgatgtttc 23700  
 tcaggcgcat gaactcgagc gcgatgatct tccatcgat cgaattgcgg gcgatgttca 23760  
 tgcccgagccg gttctcgacg atgcgctgct actgaaccag acggaagcgc atttccgcaa 23820  
 cgtgatggcg gcgaaatcgc acggtgcctg gaacctgcac ttgctcaccg cgcactgccc 23880  
 gctcgatcat ttctgctctc tctctccgcg tcgaggactg ctgggcggcg ccgccacggg 23940  
 aaactacggc gcccggaacg cctttcttga cgcgctggcc tactaccgga aggcccaagg 24000  
 ctgcggcgcg ctgagcatcg gttggggtgc gtggtcggag gtccggctgg ctgccgcgca 24060  
 ggacaatcgc ggaatcgggc tggctttgcy cggcatggaa aacctgacgc cgcaacacgg 24120  
 cctcgctatt ctggaacagc tgcgtgaacg ctccggttgc cagctgcggc cgatgcccat 24180  
 caatgtccgc cagtgccggc agttctatcc caaggcgcg cagtctgcac tgtctgagct 24240  
 tttgatgac gacgcggcga gcgaagccga tgcgccaaac cgtgtgcgcy cgcggctgca 24300  
 atcgcccgag cctcagaccc gcaggacatt gctcgaagaa catctacagc agcagctggc 24360  
 gcgcgtgctg cgcactgact ctcaaaactat cgatccctcg cgcccgctga aggaactcgg 24420  
 cttcgattct ctcactggcc tggagtttgc caacgctctc gaactcacac tgggtctcac 24480  
 gctcccccgc accctgattt ggggtcatcc cagctggccc ggtcttgccc cgcactggc 24540  
 gtgcgaaatg ggaactgccg tggctcgaagc gcaggcgcg gctgctcgcg aaggagacag 24600  
 ccgcgcatct aaaactgcac tcagcgggtt ggacgacatg tcggaagaag cagcctggcg 24660  
 tgcgctcga ggagcgaagt cgtgagggaa aaaattgcgc ccatgtcgct ggctcaaac 24720  
 gcgctattgg cgcggaacat gcggcaaaac atcgcaggct tcgacctggc tcacgccgaa 24780  
 cccatcgcca tcgtcgcatc ggcgtgtcgt tttccgggcy gcgcaagaa tccggacgcc 24840  
 ttctggagcg tgttgaagaa cgggtcacgc aggtgcgcgc agaccctggc 24900  
 aaactcgacc agtactactc ctccgatccc gatgtccgg gcaaggcgta tgcgcatat 24960  
 gccgccttgc tcgaaccgat tgaagggttc gatgcggaat tcttcggcat cccccccg 25020  
 gaagctctga acatggatcc gcagcagcgg ctgctcttgg aagtgtgctg ggaagcgcca 25080  
 gaggacggcg gcactctctc cggccctctg gcggcgagcg cgacccggct ctattgcggc 25140  
 tctcgccccc aggaactcgg actgtttcag tacgcgcgac ctgcccgcat cggagcttgg 25200  
 tcggggttcc gcgtggcgca tagcatgttg gccaatcgca tctctatct gctcgacctg 25260  
 cgcggtccga gcatggcggt cgatacggcc tgetctctcg cgctcgctcg cgtccatctg 25320  
 gcttgcctaa gctcgccgcg gcgcgaatgc gatcgggcat tcgcccggcg agtgaacttg 25380  
 atctcgactc ccgagggcat gatcgcttgg tcgaaggctc gcatgtttgg gcccgacgga 25440  
 cgtcgcaaga cgttcgacgc gcagccgcac ggttatgtgc gcggcgaggc ctgcggcatc 25500  
 gtgctgctga agcggtctct cgtatgcgtg gccgatggcg atgccatcgc tgcagtcac 25560  
 gcgggctcgg caatcaatca ggacggacgg agcaatggca tcacggcgcc gaatctgcag 25620  
 gcgcaaggag cggctctgca agagggcggt gccaacgcgc acatcgatcc atccacgta 25680  
 tcgttgatcg aggcgcattg caaggcgacg tcgctggcg atctctatga gatcgaggcc 25740  
 ctgcagtcgg tctacgaegc gccggactct gcgccttgc tgcgtgggtc cgtaaaagac 25800  
 aacatcgccc atctggaagg cgcggcgga gctgcggggc tgatcaaaag cgtactcgcc 25860  
 ctgcagcatc gcaaccttcc tcgcacactg cattttcgcc ggctgaatcc gaacatctca 25920  
 ctggacggca cgcgggttgc catcgccacg gaatcgtcgc cgtggaagc tggaaagcgg 25980  
 ccgctgttgg ccggcgctcag ctcgctcggt tttggaggga gcaacgcgca cgtcatcttc 26040  
 gaagaggcgc ctgcactccc tttgccgaag ccggtcacac gcccgagct tctcatctct 26100  
 tcggcgcgca ccgacgaag gctgcgcgaa ctggccggcc atctcgcgga gttctgcag 26160  
 tcgcaccgga atcgcttgct gtcgcagctt tgcttcacca gtcagggttg gcgcgacgca 26220  
 tatagtcacc gcttggcgat caccgcgcga gatgcggcag aggctgtagc ggcattggcc 26280  
 gcggcgccgc ggcgcgaagt atcgttgccg cggcgccggc caatcgcttt tctctcaacc 26340  
 ggccaggggc cgcagtaagc cggcatgggc gcagagcttt ataaaaagca gctctgtttt 26400  
 gcgcagcgcg tcgatcggtg gcgcgaattg ctccgtcccc agctcgatgt tccgctgacc 26460  
 gttctcttgt tcgagtcggt ttccgctgtg caagagacgg cgtataacca gcggcaaatg 26520  
 tttgcctcgg aatgggtctc ggtcagttc tggctgtcgc tcggcgctcg gccggaactac 26580  
 gtgctggggc acagtcctcg cgagtatgtt gcggcggtg tggccggcg ctttagcgtg 26640

gaggacggcc tgcggctgggt gaccgccagg gggcggtcgg tcaatgcgct tccccgcggc 26700  
 aaagcggtca tcgtttcacgc caatccgagc cgcctcgcgc cgctcgcgcg caaggtggca 26760  
 gtccgcgcgat cgaatcgccc ggaccgcacc gtgatctccg gcacggctcg agaaatcgcg 26820  
 gaagcgcaag atgacctgca tcgcgcgcggc gtggaaacgc gagagctgaa cgtatcgcat 26880  
 cggttccatt cgcgcgtgat ggaatcgatt ttggacaagt tcggaagcgt tcgaggtggc 26940  
 atcgcgtatc agccgctggc gatcccgctg gtgtcgaaac tcacggcgagc cgtattcgcc 27000  
 aaaggcacga cactcgacgc ccgctactgg cggcgacagt tgcgcgaaac cgtgcagttt 27060  
 gaaagcgcca tgcgaaccct ggccggaccgc gagtgcaagc tgtttctgga aatcgccccc 27120  
 catcccaacg tcaccacgct gggggcgatat tgcctgcgcc atgacggcgcg ggtctggctg 27180  
 cactccctat ctaagggaagc atcggattgg tccgtgctgc tggaaaagtct tggcggcctg 27240  
 tttaccgcgg gcgtgaatcc cgactggcgc ggtctctatg ccggggaaatc acccagccgc 27300  
 tgcgcgctgc cgacgtatcc gtttcagcgt gacaccttca gcctgagacg cgtaccgcgc 27360  
 agagagccgg cgcgcggcgg catgttgga ggcgcgctca acagcgcgct gggcgatgct 27420  
 atcttcgaaa attcgctaac cacggagagc cctctgctcc atgagcagct gactacgacg 27480  
 cgggtcattg tcccgcggcg ctggcacgtg tcggcatttc tcgaaagcgc acaggaagtc 27540  
 ttcggctcgg ttcctgcgc cgtctccgat gtcgatgacg ggcaggcact gggccatccc 27600  
 ccggatacgc cggtcacggt gcaagcgatt gtcaacccg gcgaggaacg cgaagcaaaag 27660  
 gtgcaggtct tcagccgatt tggcgattcg tggaaagctcc acacggcagc cgtctcgccg 27720  
 gcggcgactc ccggcgccgt tcaatccgag ctgcggcgcg agccttcgca agtcatctcc 27780  
 ggggatcgct tctacggcgc gatgaacgca cgcggcgctg atcttggccc cgccttcagt 27840  
 tgggtggaag aagtttcggc tcgcgatcgg ggcgaatgc ttgcgcgggt 27900  
 gctgaggatg gcgcgaagcg ttacccggctg caccgcggcc tgatcgattc ttgttttcaa 27960  
 gtattcgagg cgacttggcc cgcggagcgt tgcacgccc gcgcatacgt cgcggtcggg 28020  
 atcgaagcgg tgcgcttcta ccgtccgcgc gcaggttctc tgcgctgtca tgcgcgctcg 28080  
 cgcccagcgt cgagcggccc gttcgtcggt gatctgagc tggttggaag gaccggcgcg 28140  
 gtcatcgact agttttccgg actggctgta atgcatgcg gtacgtcgca atccgcacag 28200  
 tctggctcgc aggatgtgca gtggcaggag tgcgagcgat cgacaacgtt gaagtcggac 28260  
 gggcctggca agcggagga ctggttgcgt tgtgcggcg cagacgatgt cgcgggtttg 28320  
 atgcgcgaag agctgcgctg cgtgtccggc gtcaacttcc gccaggcgct ggaacagacc 28380  
 cagactttgg tcggccgccc ggcgcggctc tggctgatca cgcgcggcgt gcatcgcatc 28440  
 agtgatgagc atgcgactcc cgtcgactct ttccaggctc cactgtgggg actcgggcaag 28500  
 gcgatcgccg gcgagcatcc cgagctgtgg ggcggcctga tcgacctcgg ttgcgacaat 28560  
 gccgacatcg ccgcgcccat gctgctggat gaaatccgt atgcggcgca cgacaagaagc 28620  
 atcgcatctg gcaacggagc ccgctacgct cgcggcgctg tgcggcacaa ggaacgctcg 28680  
 aagcggccgc ctgccatttc agcccgacgc gtctatctga tcaccggcgg tctcggcgca 28740  
 ttaggacgaa ggggtggcag ccgcttgatc gagcaaggcg cgcgcgctct ggtactggct 28800  
 ggccggcata gcgaggcagt tgcgactctc gagcaactcg gggctgcagt catggttgcct 28860  
 gcttgagatg tgagtttcga gcaacagctg cgcggcgctg tggcgggacc gcgcacccag 28920  
 ccgctcgctg gactcgtgca tgcgcaggc gtgctcgatg acgggttagt tacagaacag 28980  
 accgtgggct gtttcagaaa ggtgctggcg ccgaagctgc aggggtgctg gaatcttcac 29040  
 cagcttccct gccaccatgc gctcactctt ttctgactct tctcttcgcg cgtctcgctg 29100  
 ctcggttacc cgcgacagag caattactcg cggcgcaacg catcttctcg cagccttgcc 29160  
 cacatcgccc gcgcgcaagg actaccgcgc ctgagcatca attggggacc atggcgcggc 29220  
 gaaggcatgg ccgcgcgat ccgcgcggcaa ggccctcggc gggatccgct gctgcgcgcg 29280  
 gaagtgggtg cgcgcatctt cgcgcatctc ctggcgcgaga ctggcgctca gatcggcgtg 29340  
 ttccaagtct ccgcgaaaaa aaggcggaag ccggcgagcg atcccgctt catccagcaa 29400  
 ctcaaccgag ctgcgcggga gcggcgcgag gaactgctgc agatcgcatc cgcgaagcag 29460  
 gccgcggcgg tgctggcgct cgatgcgtcc aagacgctcg acccgcgccg gccgctcaag 29520  
 gaatacggac tcgattcgct gatggcgctg gatctggcgc gcgccatcgg agagctgggt 29580  
 cgcaatgagc ttcccgcgac attgctatac gaccatccga ccgtcgagaa attggcgcgc 29640

|             |             |             |             |             |             |       |
|-------------|-------------|-------------|-------------|-------------|-------------|-------|
| catgtcctcc  | gcgaactcgg  | actcgacgtc  | cccagcgatt  | ccctcgtcga  | tgaagtgcgg  | 29700 |
| cagctgtccg  | agcaggagat  | ggcggcggtt  | atcacggaaa  | ccttgcaacca | tctgggagag  | 29760 |
| gaacgatgag  | cgatctcact  | cctcttcaac  | aggcggtcct  | ggcgctcaag  | cgcacgcgag  | 29820 |
| cgcgtctcga  | cgaaactggag | agcgtccaca  | acgaaccat   | cgcgatcgtc  | ggcatggctt  | 29880 |
| gccgctttcc  | cggcgccggac | tgcgcggaag  | cattttggca  | gctcctgcac  | gatggcatcg  | 29940 |
| atgccatccg  | cgaaaattcct | gcggggcggt  | gggatgccga  | tgcgttttac  | gatcccgatc  | 30000 |
| ccaacgcgcc  | gggaagaatg  | tacacgcgtc  | tgggcggatt  | cctcgatggt  | gccgtcgacg  | 30060 |
| gcttcgacgc  | cggctctctc  | ggaatcacgc  | cgcgcgaggt  | cgcgcgctcg  | gatccgcagc  | 30120 |
| agcgctgctc  | gctcgagggt  | gcatgggaag  | ctttggagcg  | tgcgggtcgg  | ccgcccgcga  | 30180 |
| gtctcgcggg  | cagcgacacc  | ggagtgttca  | tccggatcag  | caccgcagac  | tacagccggc  | 30240 |
| tgaaacctac  | cgatccggcg  | ctcatatga   | cctataccgg  | tacoggaaac  | gcgttcagca  | 30300 |
| ctgccgcggg  | acggatctcc  | tatctcgtcg  | ggttgccagg  | accgaacttc  | cccgctcgca  | 30360 |
| cggcggtgct  | ttcctcactc  | gtggcggttc  | atctggcgctg | ccgcagcttg  | cagtcgcgag  | 30420 |
| agtgcagcat  | ggcgctggcc  | ggcgcggtga  | acctgatctc  | ggcgccggaa  | agcacgatct  | 30480 |
| actcttgccc  | ctcgcggggc  | atggcggcgg  | atggcgggtg  | caaaagtttc  | gctgctctcg  | 30540 |
| ccgacgggta  | cggccgcggc  | gagggatgcg  | gaatgctggg  | gctgaagcgg  | ctgtcccgatg | 30600 |
| gcagcgcgtga | cggcgatcgt  | atctggcgcg  | tgattccggc  | atcggccgtc  | aaccacggcg  | 30660 |
| gccgcgcgca  | cggcctcacg  | gcgcgcgaacg | gtccggcgca  | ggaagccgtg  | attccggcgg  | 30720 |
| cgcctcaaga  | cgcgggcatg  | gccccgcggc  | atgtcgatta  | cgtggaagcc  | cacggaaccg  | 30780 |
| ggaacgcgct  | gggagatccc  | atcgaactgc  | gggcgatggc  | agcgggtgctg | ggcgaggggg  | 30840 |
| gtgcgctcga  | ttctccgttg  | atcgtcgggt  | cggtgaaaaac | caacttcggc  | cacctggagg  | 30900 |
| cggcgggcag  | tatcgcggcg  | ctgatcaaga  | ccattctcgc  | cctcgagcac  | cgagagattc  | 30960 |
| gcgcccatct  | gcatttcaac  | gcgcccaccc  | cgcacgtact  | ctggaatgag  | ctgcgcctaa  | 31020 |
| agatagccac  | cgcattgttcg | ccatggccct  | ccaacggcgcg | cccccgagtt  | cgcgggggtga | 31080 |
| gctcgttcgg  | aatcagtggt  | accaattcgc  | acgtcgtcct  | cgcgaagcgc  | aagacgaatg  | 31140 |
| tagaagcoga  | gacgaatga   | gaggcggaag  | cgaatgtaga  | ggcgaagacg  | agtaagagcg  | 31200 |
| tcaaggcgag  | tgtagaggcc  | aaagggaatg  | tggaggctaa  | ggctagtgtc  | agtgtccccc  | 31260 |
| tcctcgaggg  | ggacagccgc  | ccgcgaagcg  | gcggcggggg  | gtcggggcgg  | ccgcccagcc  | 31320 |
| gcgaggaagt  | gcgggtcccg  | gatcaactcc  | atgccgaaga  | cggccgcgaa  | tacctcctac  | 31380 |
| cgcctttcggc | gcgccatccg  | caggctctcg  | gcgatctcgc  | cggcgccctat | cgcgatgggc  | 31440 |
| gctttcacgc  | tcgcctctcc  | gcgctgtgtt  | ccgcccgcag  | ctgaocgcg   | agtcactacg  | 31500 |
| aacatcgcgc  | agcgtttgtg  | gcctcatccc  | tgcgccagtt  | caatcaattg  | ctcgaggcct  | 31560 |
| tcggcgccaa  | tgaacaacct  | cgcggcgctc  | ccaccgggtt  | cgcgcgctcc  | ggagtctcgt  | 31620 |
| cgaaactcgc  | cttcatcttt  | tcgggccagg  | gcggacagta  | cccgcgatct  | gcgtatcgcc  | 31680 |
| tgtattccga  | cgagcctgtc  | ttccgatcgg  | cgatcgaacg  | ttgcgacgcc  | gccttcgcga  | 31740 |
| gcttcgtgga  | atggcggcct  | gcggacctgc  | tcgcccagca  | gtcgggagca  | tggctgagcc  | 31800 |
| agatcgatcg  | cgtgcagcct  | gcgctgttgc  | cogttcaaat  | cgcgctggct  | gaactgctgc  | 31860 |
| aatccttcgg  | aatctgcgcc  | gacggcggtg  | ccggacacag  | catgggagaa  | gtggcgggcg  | 31920 |
| ccatctgcgc  | aggcatcttc  | accctggagg  | acgcggcccg  | catcatctgt  | cgcgcgaccc  | 31980 |
| ggctgttgtc  | cggacttcgc  | ggccggggag  | cgatggctct  | ggtcgaactg  | ccctcgctac  | 32040 |
| ggggcgaaacc | cgtctcgtcg  | gaaacgcggt  | tcactactgt  | ttctctcgcg  | ccgacgaacg  | 32100 |
| gaccacgcag  | cacgggtgtc  | tcgggagacc  | gtgtggtctc  | cgagactctg  | aaggacgact  | 32160 |
| tcgagaggcg  | cggcgctctc  | tgcgcggtga  | ttcaggtgga  | tgtcgcttca  | cacagctcgc  | 32220 |
| aggtggaaacc | cgtcgagaac  | gaattgcgcc  | aggaactcgg  | ccgcgttatt  | gcaaaaagctt | 32280 |
| ccgcgctgcc  | gtctctctcc  | acggttgaag  | gacagttgag  | cacgggcgag  | cgctgcgacg  | 32340 |
| cgtcgtactg  | ggtagccaat  | ctgcgacagc  | cagtcgggtt  | ctgggagctg  | ttgcaggcga  | 32400 |
| tggctgggtg  | tgaagttcac  | cagttcctgg  | agatcagctc  | gcactctctg  | ctgacgcgct  | 32460 |
| cgatcgaggga | tagctctcgg  | acgctgggca  | taaacgggag  | ggttcgcccc  | gtactcgcgc  | 32520 |
| gcgacgaacc  | ggagcggcgt  | gagctgctcg  | agttgctcgc  | cgcgctctac  | gtgaatgggc  | 32580 |
| acgctccgga  | ctggcgcgcg  | ctcgcctcgt  | ctcccgacac  | gcgcctggat  | ctgccgacgt  | 32640 |

```

atccctggca ggcgagcgc ttctggttcg cgacctcgac gcggcggaagt ttgccggcag 32700
ttggcggtca tccgctctct ggtcgcaagg tcgagattgc gctggcgccg gacacacacg 32760
ttctgggagtc cgtgctctct ctggatgcgc tgcggtttct gcgcgatcac cggctcaacg 32820
agcttctggt gcttcccggt gccgcttatg tggagatggc gctggcgcca gccaaaggaag 32880
tgttcgccgg tggctgcagc ctggaagaga tccggtttga acaaatgctg gttgttcctt 32940
ccgcggggcgc ctccgagtg caggtcatac tcgaggggaca cgcattccgc atctccagtc 33000
tggccgaagg cggttccgat tggaccgagc acgcgcgcgg caccattggct gggcgcccg 33060
acaaggctgc gccccaggtg agcctgccca cacttgggga tcgcacgcag ggcgatgact 33120
tctatgcggc cttcgcatcg caggggatgc attacggcga caccctccgc ggcattccgc 33180
aagtctggcg gcgcgacggc gaggcagtg cgcgactgag cgtgccggat gccgttcgcg 33240
aagcagagtc cggttacaog cttcatcctg ccttgcctga tgctgtttg cagggtgctgg 33300
gcgcgacgct tggcgccgaa ggcagcgcgc gtccttgctg gctgtgcgc atcgaaacgg 33360
tgcactgttt cggcagaccc gccggcgatc ttagggtgca tgccggtcg acggggcgcc 33420
tcgagggcgca tgtcaccttg tgtgatgcgg aaggccacgt catcctcgag gtccaaggcc 33480
tgctgtccca ggaactggag cgccaatcga aatgggtcca cgctatggaa tgggagccgc 33540
agctcgtggc cgagagttca acggcaacgg tgcgggtgct atggctggtc attgccgatg 33600
ccggcgccat cgcagccgcg gtggcgcgag ggctggggca aaacacgggt ggtgattccg 33660
gtcgcgatcg cgagataccc gatcagcctt accggggcgt cattactgc gggagcctgg 33720
atgagaccga ggatgagacc gatccgtcgg ctgcgggggg aaccgcctgc gaagacattt 33780
tgccgatcgt tcaagaattc ggaatgggac gcatacagc gacgaacaa gcgtccgacg 33840
ccgaatcgca gcattccgca atctggctga ttacggcggg cgttcacgc gagcatctgc 33900
agatgccggt ggtgcccgcg cgggcacccg tgtggggctc gggacgtacc atcgcgccg 33960
agcatccgga gttgcgttgc acctgcacg atctcgacac tgcccggtgaa gcgcaggtgc 34020
aggcgctctg ccgagagatt ctccggggga gttctgaacg tcagggcccg g 34071

```

&lt;210&gt; 115

&lt;211&gt; 4615

&lt;212&gt; ADN

&lt;213&gt; bacterie

&lt;400&gt; 115

```

actgcagtcg ccggaatcgg cgggtggaact acagcagccg ctggtgcgta tgggattgga 60
ctcgctcatg gcggtcgcaat tacgaacccg gatctgcgcg tcttgcctcc 120
catggtccga tttctagacg gccccagcgt tgggaaactg gccagggatc taagcatct 180
aagcgccctc agcgaacgca cgacggtggc gccggaacct gcggcgccag cctcggttcc 240
tgccctctcc tacctctca gcgcggcgca gcaggcgctt tggtttattt accgaagcgc 300
gcgggaagtg cccgcataca acatcgctg gatcgccgcg gcgagaggcg ctttcgatcc 360
gcaggcggtg gcgcgttcgc tgcaaggact ggtggatcgt catccggcgc tgcgaacgac 420
gattggcgag agtggcgggc caccggttca aacgggtccac agcagcgtcc cgttgatttt 480
cgaagtgcgt ccgtgttcgc cggacgatga ggcggtgctg atcgacggcg ctttcacgcg 540
gcccttcaat ctccgcgaaa actgtttccg ctccgctctc ctggtgactg ccgggaagga 600
tcagggttct gccatcgtgg tgcatcacat cctcgccgac ttctggtcac tgctggtgat 660
ggtggatgaa ctccgcagta tctacctcgc gaggacagct ggcggtccgc ctgtcgccgc 720
gccggtccgc agcttcgcgc ctttcgtccg ctggcagaac gaactgttgg ccggaaccga 780
gggcgagcgg ctttggaaact actggtcctc gcagctttcc gccagcttcc cgtttctgaa 840
tctcccgctg gatcgtccca gtccgcgggt gcagagtttc cggggaaact ctcaactggt 900
ccgaatcgaa cccgcgctga ctgcgaaact gaaggcgctc gcgcggcgac agaacgcgac 960
gctgcacgcg acgctgatgg cggcgtttca agtgcttctc tcccgctgga cctcacaaga 1020
agagatcctg accggcacc ccaaccaagg tcggacgcaa ccggaattcg ccgatctcgt 1080

```

cggataacttc gtgaatcccg taatccctgcg aggagaactt tcaggcgatc cggatttcaa 1140  
 tacggtgctc gcccggtatc ggcaaacgct tctcggcgcg atcggagcacc agggagtacc 1200  
 gtatgccccg atcgtggagc ggttgggtcc cggactcgcg gttctattcg tgctccagca 1260  
 gcctcatcgc attccccgaat ccgtgcgcgt catgttgggt cagtcggcg gtcgcattggc 1320  
 ctggggcagc ctccactcgg agtccctggc gatgccgctg cgacagagcc gggttgacct 1380  
 ggatctgatg atggtcgaaa ccgatggagg cctctccgcc tttctgcaat acaacacgga 1440  
 catttttgat gctgccacga ttgaacgtct ctcttgacac ttcgccgtgc tgctggaagg 1500  
 aatcgcggag aatccccctc gtccagttgt cgatctaccg ctgctgacaa ccgcgggaacg 1560  
 catccagctg ctcgaaagagt ggaatgcgac cgccgcggaa ttcgcgtccc aatcgtgca 1620  
 cgagctgttc gaagctcagg tggagttgac gcccgacgcc atcgcgttga gcttcggtga 1680  
 gcagaatctg acatatcgcg aactcaacgg gagcgccaac cggatcgcg actatctccg 1740  
 ctcgccggcg gctggaccgg gcgaaatggt tggcatccat gtcacgcggt cgctcgaaac 1800  
 cgctcgacgg ctggtggggc tccctgaaggc cggcgcgccc tacgttccgc tggaaacgga 1860  
 atatccggcg caacgtcttc ggtctgatct ggaagagacc aggcggcgct tbtgctgaa 1920  
 tgtcacggaa tcggaagtat ggaacgcagg cgacaccaat ccgaacccgc tgcgcactcc 1980  
 cgccgatctc gcttatgtcc tgtacacct ccggttcgacc ggccggccga aaggcgttga 2040  
 aatcacacac caggccgctg tcaattttct ttcgtcgatg cggcatgagc cgggcattag 2100  
 cgaccgcgac agcgtgctcg ccctcacgac gttcatgttc gacatttcgc cgctcgagat 2160  
 ctttttgccc ttgagcgccg gcgcgcgcgt cgtggtggcg aaccaggaga cggccgtcga 2220  
 tggtagaggg ctggcgaggg aactcgcggc cagcaaaagc acaatgatgc aggaactcc 2280  
 cgccactcgg cgctcgtcgc tcgcacccgc ctggcccgcg gacgcgcgct tgcacgcgct 2340  
 ctgcgcgggt gaagcccttc ctgcgcgact tgcgcacggc ctctgcaac gaacgcggcg 2400  
 gcatggaat ctttacggac ctaccgaaac gacaatttgg tcgcgcaccc aacgggtgac 2460  
 gacaggtgac ggaccgggtt cgattggcgc ccccatcgca aacactcagc tctatgtgct 2520  
 tgacgatcgc atcgaccggc caccatctcg tgttgcgggc gaactgtaca tggcgcgcg 2580  
 cggcgctccc cgtggatacc tgaatcgctc ggaactcagc gcggacaagt tcgtcgccaa 2640  
 tctgttcgac cctcatggca ctgcggtgta tcgcacggga gatctcgccc gcgcgcaacg 2700  
 cgacggcgcg ctcgagtatc tcggccggat cgaccaccag gtgaagatag cggggttccg 2760  
 catcgaaacc ggcgagatcg aggcgcgcgt ccgcagtcac ccggcggtcc gatcgtgt 2820  
 ggtcacccgc agagaaaaat acgcggccgg taagtattct gcggctcaca tbtcccccct 2880  
 tgtcgacggc ctgcgcgca cggcagccgc cgacacattc cagcaccgag tcgagtcga 2940  
 gcacgtgacg cagtggcaat ccgtctggga caccacatat gaacgaatg ccgcgaacgc 3000  
 ggaatccggag ttcaacatcg tcggctggag aagcagtggt accggagagc cgattccaag 3060  
 tgcccgagat cgggagtggt tgcaggattc cgtcgatcgc atctcgccct cgccgcgcg 3120  
 tcgcgtgctc gagattggct gtggtacggg actgctgctc ttcgcgctcg ctccccactg 3180  
 ttcggagtat tggggcacgg acttttcgca gaaggcgctg gactacatcg ccgctcacgc 3240  
 ggaccgcacc ggctgggcaa atgtccgcac gttccggcag gcggctgcca acgcttgca 3300  
 gatcgacagt cgctcgtgcg atcgggttgt tctgaactcc gttatccagt acttccccg 3360  
 cgaagcgtat ctgcggcgcg tgcgtggcga ggcggtgcgt gtggtcaaac cggggcgcat 3420  
 cgattttgtc ggcgattgct gcagctctccc gctgctggag acgttttaag cttctttag 3480  
 agttcagcgg gcacccgcgt cgttggacgg gaatgagttt cggcaacgcg tctgcttgc 3540  
 cgcgctcgag gaagaggaac tctgtgtcga tccgcgcttc ttttttgctc tccgcgaaca 3600  
 gattccggag atcggccgga ttgaaatcct gccgcgtcgc ggcgggtcgc ataacgagct 3660  
 gaccgccttc cgctaccagg cgatcctgca tatcggtcg cgggaagcgg aggagccgga 3720  
 atcgagctcg aggcgttgcc agaccgcggc cgaaatacgc agagatacga cggacgctca 3780  
 gccggagtgt gccgcattta ccgagattcc gaacgcacgg ttgaccgcgc aaagcgcat 3840  
 tgtgactgag atgaacggtg acgaagctcc agagacactc ggggagttgc gggaccggct 3900  
 gcgcacggcg tgcctctccg gcgtcgatcc cgcgcatatc tggcgatatg acgaagacct 3960  
 gccgtacgcg gtggcaatcg actggagcag tcatgggcca cagcgacgct tcgacgcgac 4020  
 cttctgcggt gcggcgggcg gtcgcgcggc ttcgcgtccg cgaagccgcc tggccggccc 4080

gtatacgaac gatcogctgc gagccgtcta tacgcgcacg gttgtgccgc agttgcgtac 4140  
 tcaatctgaag gagaagctgc ccgactacat gatccccgacc gcgtgggtgc tgctccacga 4200  
 aatgcgcgtgc acgccccaacg gaaaaatcga ccgtaacgcc ctgcccgcgc ccgagcccag 4260  
 ccggcgagcc caccgccaag cattcaacgcc tccggaaact ccggtggaac aggtactcgc 4320  
 cacacattgg ggcgaggtgc tcggcatgga tggcatcgcc gtccatgac actctcttga 4380  
 ctctggagga cattcgtcgc tggtcacgca gatgatcgcc cgctgctgcg acatgctcca 4440  
 cgtggaagtgc cccttcgaa ccgtgtttaa cgcgccacg gttcgaggtc tcgcccgcgc 4500  
 tattcaggac ggcgtagacc caggatgggc aaggcgagcc gccgatttgc tgatcgctgt 4560  
 ttcccaaatg tcagatgttc aaatcgagcg tatgatgagc gccgcccacg actag 4615

&lt;210&gt; 116

&lt;211&gt; 8301

&lt;212&gt; ADN

&lt;213&gt; bacterie

&lt;400&gt; 116

atgcagaatt cgtcgccaaa taccatagac ctctcgctcg cccgcgcgca attgctcgac 60  
 cgtctcgtgc aggaaaacag cccgaacat cgcacccgc gcgtgaaaa ccgggatgcc 120  
 gcacccctgt cgtcggccca gcagcggtt tgggttctcc atcagctcga cccgattctc 180  
 ccgccttaca acattcccat agcgtcgc atccgaggtc cgctggatat tgcgctctcc 240  
 ctgoggagtc tggaggcgtt ggtgcagcgc caccagagcc tgcgcagctc cattggcgtt 300  
 gtggatggag aggcgcgcca gagcctcctg gcgcgagtga cactggaact tccggttgtt 360  
 caggctgacg gaatcgcaga agcgcggcaa atggccttgc gtgatcccca gatcccgctc 420  
 gacctgcgaa aaccccgcgt tctgoggacc aagctgatct gcctcgatga caagcagcac 480  
 attctcctgc tgacgttgtg ccacatcatc gcggatgctt ggtcgtcga gacgttcgtc 540  
 cgcgacctga cgcgatcgta cgaagcgttc gtgcaggggc ggccatcgcc gctcatggaa 600  
 ctgcgattc agtatggcga ctgggcgcgc catcagcaga cgtcgtgaa ccaaacccgc 660  
 cagcagtaac ggaagaaaca gctgtcgggc acctgcctt tctcgcacct tctaccgat 720  
 cgcccccgcc cgcgcgagca gacctggcgg ggccgcgtgc agaccacagc cctcgccgtt 780  
 gatttgacgc atggactcca cgcgtttgcc ttgcgtgaa gcgcgacggt gttcatgacg 840  
 gcaatcgccg cgtttcaggt gctgctgc atcctatacc gcaggaaga catccttacc 900  
 ggggttccag tcgcgggcgc tacacaacga gaacgggaag gtctcgtcgg ttgtttcgcc 960  
 aacatgagc tctcgcggcg cgaatcgccg gcgatccgt cgtttcgagc tcttctcgcc 1020  
 cgcacccgcg acaccgcctt gagcgcctc tctcatcagg actttcctt cgaacgcctg 1080  
 gttgaggaa tcgatcctcc gcgggacctg agccggtcgc ctgtatttca ggtctcctc 1140  
 gcgctgcgc ccgatgcgcc ggcacaccc gtcatgcctg ggctcaccat ctgcgcgag 1200  
 tacatgcaca acgcgcgatc caaactcgac ctgcgcgtga cctcgcagcc atccggcgat 1260  
 ggactgatgg cgtccgcgca atacaacacc gattgttg agttgttcg atgcggcaac catcgctccc 1320  
 ctgctcgatg cgtaccgaac cctgctggcg agcgtgggtga ccgattccga cgtccgcatt 1380  
 tcaaccgcgt cgtgttgttc cccgcgggtc cgaagccgga tgctcgagca gcaaatgcy 1440  
 acacgcgcgc atgcgggtcc gaacgggtgt gcgatgaac tggctgaagc tcaggccgaa 1500  
 cgcactccgc acgcccgcgc cgttgtcttc gaagaccatc agttgaacta cgcgcagctg 1560  
 aatgcgcggg ccaaccgcct ggctcatcgt tagagcgcat ccggcgccgg ccggggaag 1620  
 atcatgcctc tggcgatgga gcgctcgctg cgtatgggtga ttgcgctgct tgcgattctg 1680  
 aagtccggca gcgcgtacct gcctctcgat cccgcgcacc ccaaggatcg tctcgcggcg 1740  
 attctcgatg aagtgcaccc gcacgcggtc ctacgcagg aggcgggtgc tgagatgatg 1800  
 gcgatgatgg cgatgatggc ggtcgcgcgc gaaccagaag ctgcgaaatc cgtcagcgcg 1860  
 agcaagcccg acgatctcgc ctacatcata tatacctcgc gatcgacggg gcgacccaag 1920  
 ggcgtggaga tccgccatgc tgcgtagtgc aactcgtcgc gctccatgca gcgcgagccg 1980

ggtctgacag ccgcccagtg gctggtcgcc gtcaccaccg tgtcattcga tattgccgga 2040  
 ctggagatct ggctgcggtt gatcacccgc gcccgcgcca tegtgcgccac ccgcgagatc 2100  
 gtgggttgacg gcgagcggct caccaccctg ctggataagt cgggcgctac ggtcatcgag 2160  
 gcgaccccca gcggttgccg gcaattgctg gattccgggtt ggaagccggg taaaggcttc 2220  
 cgtgttttct gcggcggtga agctctcgct ccggaactgg cgcgcgcgat tctcgatagt 2280  
 ggcgtagagc tgtggaatct ttaccggacc caggagacca ccatatggct ggccgtgcac 2340  
 aagacacaaa gactgggtgc ctccgatagc atcgtgccga tccggccatc catcgacaa 2400  
 acgcagttat acatcctgga ttccgcgatg gagccggttc cccccggagt tccgggagag 2460  
 ctgtacatcg gaggagcggg actggcgccg ggctatcatc gcaaccccga gctcacgcgt 2520  
 gagaatttcc gcgagtgggc tgatcgagga cgcatttact ctaccggcga tctggctcgc 2580  
 taccgttccg acggcgagct cgaagtgcct ggacgagtcg atcgccagat caagctgcgc 2640  
 gggtttcgca tcgaaccggc cgagattgag gcccgcatcg agacgcacat tgcctgtaag 2700  
 caggcgatta cggctcgtgaa ggacgatcgg ctgatcgctc atctcgttcc ggcaacgggc 2760  
 gacgtcgccg atctcgagag cgattttcgg tcgtggctgg caacgcgcct tcccgat tac 2820  
 atgatccccct cggcggtttgt cagcctgtcc tcccttccgc tgaccgcaaa cggcaaaatc 2880  
 gacgcgaacg cgtctcccg tttgccaca acgcccgttg ctgctcgca gccgatgcgc 2940  
 ggcgatgttg tggagacgat tgcgtccatc tggcgtaag ttctgcgctg ggagcagctc 3000  
 gactatcgcc agaactctct tgatgtcgcc gggaactcgc taatgctcac acgggtgcgc 3060  
 ggactgtcgc aggaagcgcct ggggttgacg ctctccgtcg tcgatctgtt ccggcatacg 3120  
 acgatcgagt cgtctgccgg cctggcgaaa aaatccgaac ccgcgcgtgc ggaacctgcy 3180  
 gctcggtcgc cagaagatcg gatcgagtt atcgggatgg ccggcggtt ccggggggcg 3240  
 gcgaatgttg aggaagtctg gcgcaatctg cgcgacggtg tggattccat cgccaggtct 3300  
 tcgcccgaag atctcgtggc gggcggcate agcccggagg tcttcaggga cccgagctac 3360  
 gtgcccggca agggctctgt ggaaggcgc gagtttttcg atgcccgtt cttcgctac 3420  
 agtcgcgcgc aagcggagat catggaccgc cagcatcgcg tgtttctcga tgcgcgctg 3480  
 gaagcgatga agaacgcggg atatgcggcg cgaagctata agggttcgtat cggcgttttc 3540  
 gcgggatgag gcgtcaatac ctacctgctg aacaacctcg ccaccggcga gccgttcat 3600  
 ttctcacgcc cctccgcgta ccagctgtcg acggccaacg acaaggattt cctggccacg 3660  
 cgtgtctctt acaagctgaa cctccgcggg ccagcctga cggttcagac ggctgtctcc 3720  
 acctcgtcgg tgtcggtggt gatggcatgc gagagcttgc agcgcggcgc ctcgacatt 3780  
 gccttgctgg ggggagttgc ctcaaatggt ccgcagtcg tggggtaact gcaaccgcg 3840  
 ggcatgatcc tgtgcccgga cgggcgctgc cgcgccttcg atgagtcgc tcaaggcacg 3900  
 gtgcggggca acggcgcggg tgtggtcgtc ctcaagcgtc tgagccgcgc tctggccgat 3960  
 ggcgacacga tctacgcgct cattcgcgga gcggctatta ataatgatg ccgcgagcgc 4020  
 atggggttta ccgctccagg tgtggacggt cagacgcgat tgattcgcg cactcaagag 4080  
 atggcggggc tgaagccgga gtccatcgcc tacatcgagg cccaccggaac agccacgcgc 4140  
 ctgcggcgct cggttgagat cgccgcgcac cgcgccaaat ttccgaaaaa cgggaagcgcg 4200  
 gatgtgtata tcggatccgt caagaccaac atcggtcatc tagacgtctc ggccggtgtg 4260  
 gccggctcgt tcaagacggt cgttgcgcgc catcgcgccc agattccgc cagcctgaa 4320  
 ttccagcgtc cgaatccgcg aattgatttc gcaaacactc cgtttcgtgt gagtacgcgg 4380  
 ctgctcgagt ggcccgcggg aaagaccgcc agacgagcgg cagtcagttc gttccgggac 4440  
 ggcgacacca acgctcacgt gatctggag caagcgccgc cggtagacc ccgcgcagct 4500  
 gcgcccgaac gatccgcaca tgtgctttgc ctgtccgcca atacagacgc ggccctcgaa 4560  
 gaactgtgct gctcgtatgc cggccatag gacaaccagc ccggtttgtc gttcggcgat 4620  
 gtccgatcca cggccaatgc agggcgcggt cacttccgc accgtatctg cattgtggcc 4680  
 cggtcgagcg acgaggtcgc ccaacgactg accggagcac gacgggttgc catcgccac 4740  
 acgcgcccca agatgtcgct tcttttcacc gcgaagggtg cgcaatacgc gggcagggc 4800  
 cgccagttct acgagtcgca gccggtgttt cgcgcgcgca tggatgaatg ccgagctctg 4860  
 ctgaatggac ggtcgatct gccgcgcgtg ttggccgatg acgctgtgct cgacgcgacc 4920  
 gccggcgcgcc agcccgcgt gtttgctttg cagtgggcct tggcgagttc gtggaagtcc 4980

tgggggtgtga cggccgacgt ggtgatggga cacagcgtcg gcgaaatcgc ggccggcgtgt 5040  
 attgcccggcg cegtgcagcct gccggatgag ctccggttag ttgcgaacg cggccggctc 5100  
 atgcagaacc tgcgggaagg tgcgatggct gcggtcagcg ccggcgagca gcgctgtgcc 5160  
 cgagcgatca cctcgcgcgt ctccatttgc gccatcaacg gaccgcgtga ggctcgtgatt 5220  
 tcgggtgcgc cgcaggaatg tgagagcgcg ctggcaactc tacgtgcgga gggcatcaaa 5280  
 acgcagatgc tggccgttgc gcgcgccttt cacagctcga gcatggaatc gattctggcg 5340  
 gacctgcaac gccggggcggc ggcgatcgcg tggcgcaatc cttcgatcgg cttggtttcg 5400  
 aacctcacgg gcaaaactggc cggcgagggga cagctggcga atccgctgta ctggcgagat 5460  
 cacgctcgaa accctgtccg tttcgccgac ggtatccaaa cgctcaagga caagggctcg 5520  
 gacggttttc tcgagatcgg tcttaagccg gttctactcg gcatggggca aaagtgcctg 5580  
 cccgacgacg ccaagcagtg gctgccgtcg ctgcgtaaag gccgcgatga gtgggagacg 5640  
 attctcagca gtgtggcgac gctatatcag ggtgggttcg acatcgattg gcaggagttc 5700  
 gaccgtccgt attcgcgaag gcgtgtcgcc ctgcgggctc atcctttcga gagacgcgcg 5760  
 cattggatcg agcggagttc cagaccggaa cctgtagcgg ttgcgagtggt tctcgtcggg 5820  
 tgccggctgt cgctacccgt ggcagaagtt atcttcgagt cgaactatc gcaggcttcg 5880  
 cctctactct cagaccacg atattacggt tcggtgggtg ccgcggcggt gtaactctcg 5940  
 gccatggcgcg tcgaggcgct ggccggaggtg tttggcgccg gccggcacac cgtggaaaa 6000  
 gtgaacttgc gtaacctct gatactttca cgcggagcgcg acacgcgtgt tcagctcgtg 6060  
 ctttcacaga gcgatgacg gcatgctcgt ttcgcatac tcagcttgct cgacggctcg 6120  
 tggaaactac atgtgtccgg caatattgcc gccacgctg gtgtcgtccc cgtgccccga 6180  
 ctggttcgat aacgcgcggc tgccgtggat ggagacacgt actattcgct cgtgcgcaac 6240  
 ctcgagatag aactggggcg gactacacgc cgcatacagc gcattcattt cgtgtaaacg 6300  
 gaagcgctgg ccgcgat tga ttcgcgaacg ccgctcaatc ccggtgtgta attggcgaa 6360  
 gccggctcgc aattgcttag gcgcgcggcg agtcccgcgc ttgcggatgg ccgccgaact 6420  
 ccatattcgt ctccgctcgg tatcgatcgc gtttgttttt accggagcgt ggagggcgcc 6480  
 gtatgggggg ccgcgcaaat tctccggcat tcgcgggacg gctttaccgg cgaggcgcgag 6540  
 ttgctggaact cggagggctg cgttctcggg gaacttcagg gcgtgagttt ccggcgcgctc 6600  
 actcgcgcat gggcgacgct ctccggaacg aagcccgaaat tgatgaggtt cgagtggcg 6660  
 cccgaaccgc tccgccagcc ttccgcaacg ctacagcctg gggcatggct gatcctggcc 6720  
 gacagtggcg gcgcggcccg cgctctggga gatgcgctca cagctcaggg cgagatgtgc 6780  
 gttaccgtgc ccgcagcctg cagagtacatg tccctagtgc gtgagcgtga ctggcgccgg 6840  
 atcgtaaac tgtacagtct cgtatgattat gagctcggct gcgcgacac ttgcggcctg 6900  
 gtgaagtccc tgaagtccgg tccgcggcta tggtggttaa cggccggcgcg gcaggcgacc 6960  
 agtgcgtgac acaattccat cgaggcgcgct ctctgggctc tcggccgggt gatcgccgcg 7020  
 gagcaccgg atctgtgggg cgggctcact gatctggatc ccgacgatgc gcatgcttcg 7080  
 gcggccggcg cggccgcgca gatgcgtgat ttcgacggcg aagatcagtc ggctgga 7140  
 agcaaccgcg gctacgtgac gcgactgacc cgcgcgaacc gcgcgcgac gcagctccgt 7200  
 ctgggtttcgg gcgcgactta tttgatcacc ggcgggctcg gagccctggg acttacagtc 7260  
 gcgaaatgga tgggtgagca cggcgccact cgctcgtgc tggccggggc ggccgctcca 7320  
 aacgaggagc agcagcgcgt gctgcaacag atgtgtgcga cggcagagac ggtcgacgtc 7380  
 agccgggaa aagaggtcgc ggtatctcatt cgcgcgaccc acaccgaaac gtacacgctg 7440  
 cgcggggtta tccatgcgcg ggggtgtgctg gacgacggcg tactcgtgaa tcagccgcg 7500  
 acgcggaatg caagcgtcat ggccgcgaag gcggaaggcg ctgtacacct ccatcatcac 7560  
 accgcgcatc tgcgcgtcga cttctcgtgt cttcttttct cggcatcctc gctcttaggt 7620  
 cctgcggggc aggcaggcta cgcgcgggct aacgcgcttc tcgattcgct gcgcatacac 7680  
 cggcgcggaac tgggtttgcc ggcgaccagc attaaactgg ggctctggtc gggagccgga 7740  
 atggccgcgc gcaccagcca gtcgatggcc ggctgggca gcctctcgtt ggacagggtt 7800  
 ctacacatc tcgaggcgct cctgcatgaa tgcaccttc agattgcgc gcagccggcg 7860  
 ggctcgatta ccggcgagtt gctgcgtccc gccgcgctgc cttcactcca actgcgac 7920  
 cgcttgaacg aagccacac ccggcagcgc gaagccatcc tcattgcgca catcagggag 7980



|             |            |            |            |            |             |      |
|-------------|------------|------------|------------|------------|-------------|------|
| tcaactggcgc | gctttgtcgg | catcgcgact | tccacaccgc | tcatccaca  | gcagcctttg  | 8040 |
| gggtgaactgg | gactcgattc | gctaattggc | atagaacttc | gcaactcgct | ctcccaatca  | 8100 |
| ctgggggcagc | ctttgcccg  | gagtcgtcgt | ttcgactatc | cgctcgctga | tgcgatcgctc | 8160 |
| agttacgtgc  | tccatgcggt | atttccacc  | gaagcatcac | cggtggaagc | gccggagttt  | 8220 |
| gagaacctcg  | ccgcggaag  | actggaagcg | ctgctcgatt | cgcggctggc | gcaggtcgac  | 8280 |
| cagtggttgg  | agacgcaata | a          |            |            |             | 8301 |

&lt;210&gt; 117

&lt;211&gt; 5292

&lt;212&gt; ADN

&lt;213&gt; bacterie

&lt;400&gt; 117

|             |             |             |             |             |             |      |
|-------------|-------------|-------------|-------------|-------------|-------------|------|
| atgagcgggt  | cagacgatct  | cagcaagctt  | cgccgcgcgc  | tgattgcgct  | cgacaagggtg | 60   |
| cagaaaacgca | tcgaccagct  | ggagagcgcg  | cgcagcgagc  | ccatcgccct  | catcggcgcg  | 120  |
| ggctgcccgt  | tccccggcgc  | atccaatctc  | gatgcctatt  | ggctgctgct  | gcgcgagggc  | 180  |
| cgcagcgcg   | taegtgaagt  | tccaccgcac  | cgctgggaca  | tcgatgccta  | ctacgatccg  | 240  |
| gatccccggc  | cgacggggcg  | aattgtacacg | cggtacggcg  | gcttcacoga  | tcaggttgac  | 300  |
| cgttttgacg  | cccggttctt  | cggcatcgct  | ccgcgcgagg  | cgatcagcct  | ggatccacag  | 360  |
| cagcggtcgc  | ttctgggaagt | cacctgggag  | gcgatcgaga  | acgcggggct  | tcaccacgca  | 420  |
| cggttggcgc  | ggagccggac  | cggcgtcttc  | atggggatct  | tttccaacga  | ttattacaac  | 480  |
| ctgcaaatgc  | cgcgccggga  | tgcgcatact  | gacgcgtaca  | cgggcacggg  | caatacgccc  | 540  |
| agcgttgccg  | ccgggcgtct  | ctcgtacatc  | ctcgggctgc  | agggcccgaa  | catggcgatc  | 600  |
| gacacggcat  | gctcgtcatc  | gctggctgcg  | gtgcaccttg  | cctgtcagag  | cctgcgctca  | 660  |
| ggtgaaagcg  | acctcgcgct  | ggcgggcgcg  | gtcaactctga | ttctctcgcc  | ggatcggaag  | 720  |
| atctactctt  | gcaagctgaa  | ggcgatggca  | gccgacgggt  | gctgtaaggc  | attcgatgcc  | 780  |
| gcagcagacg  | gctacgtccg  | cggtgagggc  | tgcggtgtgg  | ttgtgctgaa  | gcgactctcc  | 840  |
| gacgcgctgc  | gcgatcgca   | tccgggtgatg | gcgggtgattc | gcggcacggc  | aatcaaccag  | 900  |
| gacggacgca  | gcaatggact  | gacggcgccg  | aacggggccc  | cacaggaagc  | cgtgatccgc  | 960  |
| caggctgtgg  | gagacgcgcg  | cttgacagacg | ctggatgtga  | gctatgtcga  | ggcgacggga  | 1020 |
| accggcacgc  | cgctggggca  | tcccatcgaa  | gccggagccc  | ttgcggcgcg  | gctgggagcg  | 1080 |
| gggcgcacca  | acggcaacaa  | gctgaagctc  | gggtcggtga  | agaccaactt  | cgggcacctc  | 1140 |
| cgggcgcgca  | cgggcggtgc  | cgactgatc   | aaggtggcgc  | tgatgctgca  | gaacgaagcc  | 1200 |
| attcgcggcc  | atctgaatct  | gaccacggcc  | agcccgacac  | tcgattgga   | cacgcttccc  | 1260 |
| ctcgaaatcc  | cggcacggct  | caccgccctg  | cgggttgac   | ccggcgggcg  | gcgcgtcgcc  | 1320 |
| ggcatcaact  | cgttcggctt  | gagcggatag  | aatgcgcacg  | tgctcatcga  | gcaggcgccg  | 1380 |
| caacaggcgc  | cgtccagtag  | gcccgcacgc  | tacctgtctc  | cgctatcggc  | gcgcagtcgc  | 1440 |
| gaggcgcctg  | gtgatctggc  | gcgcgcatac  | gcgcagctgg  | tgaacgacaa  | ccccgcgcag  | 1500 |
| acctgctaca  | cggcgtgcgc  | tcgcgcgact  | tcatacgaa   | accgcgcggc  | attcacgggg  | 1560 |
| acgaacgcgc  | aggacttgat  | ggccgggctg  | gacagttttc  | tgccggggcaa | cccgaaccgc  | 1620 |
| gataccgcga  | cagggtttgt  | gcgcgcggcg  | gacaagcgaa  | aagtcgtttt  | cgctttggcg  | 1680 |
| ggacaaggat  | cgagtgggc   | cggcatgggc  | cgacactcga  | tggtctctga  | accggtgttc  | 1740 |
| cgtgcgcgca  | tcgaagagtg  | cggccgcgcg  | atgcagcctt  | acgtcgactg  | gtcgcgtgac  | 1800 |
| caagagttgc  | agggggcgct  | cgaccgcatac | gacgtgattc  | aaccggccct  | gttcgcagtc  | 1860 |
| ggggtcgctc  | tgccgggact  | gtggcgccat  | tggggaatcg  | agccggagcg  | cgtgatcgcc  | 1920 |
| cacagctagg  | gcgaagtgcg  | ggcagcgcac  | atgtcaggtg  | cgtgactct   | cgatgaagcc  | 1980 |
| gctcgggtga  | tttgctcgctg | cagcccgatg  | ctgcgcggag  | tacgcggcca  | ggggaagaatg | 2040 |
| gctgtcgtgg  | aattagcgct  | ggacgaggcc  | atcgctgcca  | tcggcgggcg  | ctcggtatcg  | 2100 |
| gtctcgattg  | ccgccagcaa  | cagcccgcg   | agcacgctcc  | tgtcgggcga  | cagcgacgct  | 2160 |

ctggcgcaac tgcctcgcca actggaggcg aaagacgtct tctgccgtcg cgtgaaagtg 2220  
 gacatttccct cgcacagcca tctgatggac tccgtgtgcg cggcgttgcc gggcgtgggtg 2280  
 ggagcgtctc agcccgccgc ggcgcgcctt ggcattgtact ccaccgtcac cggcgcagcg 2340  
 attagcgggt aagagctgggt ttctgcgtac tgggctcgta atcttcgcca acccgctgatg 2400  
 ctgtcgacgg cgtgcgcccg agccgcggcg ggtggctatg atgtgtttct ggaactgagt 2460  
 ccccaccctg tgttggtcca gccgatccag gaaacgctcg gagatcgggc agcgattgcc 2520  
 gctgctcgtg tgcggcgcca tgaagacgga aacctgcac tgcgcgggac gctggggagcg 2580  
 ctgctgacta accgagtcac tccggactgg tctcgtattt atcccaacgg cggccaaact 2640  
 gcgcggctgc ccaactatcc ctggcagcgt gagcgttatt ggcattgatc cgttccgcgc 2700  
 cagggtcagt ctcaggcttt gccctggccgg cggatcccg tgcgcgtgcc ggagatgcag 2760  
 ttccagtcga ctgtggaggg gaaagatttc gcggatcacc ggcgtgcaca tgtgatcgtg 2820  
 actccggggg cgtggcactt ggcaatggcg ctccgcgtcg cgcgccaaag tctcggcgcc 2880  
 gggcctcacc atgtcgaaaca cgtgtcattg acggcgcgcg tgacgtgcc ggaaaacgat 2940  
 gctgccaggg aggttcaact ggtactccgt catgaagagg cggcgggagc ttcttccgc 3000  
 atctacagcc gcgagatttc ctggaagctt cacagcgaa gcatgctgca ggcggggcag 3060  
 tccaccgcat ccatcgatct ggcattgcgt cgcgcgcgtt gcaaggcgga gctcacagcc 3120  
 gatgccttct attccgcatc gtgggactgc ggcctacact tccgtcccaac ctccgaaac 3180  
 atcgccacca ctggcgcggg caacgggtgag gtgctttgtc gcgtgacac ttccgtgacg 3240  
 gaaatgcaga cgcctcactg ctgtctgcag ttgcccgcgg cctcgttcca tcacgacgat 3300  
 ttgaaagatt tgcattgtcc ggtaggctcg gaccgatttc cgtcgtcgta agtgcccaat 3360  
 gggcggctgt ggggatacgc ggtcttgcgg ccggattcca cgtgtcgttc 3420  
 accggcagcg gcagcgtggt gggcgaaatt gtggggctcg agtcgagagt cgcctaatgc 3480  
 ggccagctcg cgaatcgga gattccacc tggacgggtc aatggaccgc gtcggttcgc 3540  
 cgcggcgatg ccaatgcggg caatgctggc ggaccttggc tegtcatcgg cgagccggcg 3600  
 attgcgagga ctctgcaaaa gcgcggccaa acctgcgcga cggccgatac gtgctcgggt 3660  
 cgcgcgtgcc gtcaaatgtg tctactgtcc tgcgcgcga tcgacgacct gctttccgta 3720  
 ttgcgcagca tctgtcaagc gggctggcct gagccgcgc gctgtggct gctgacgcgc 3780  
 ggcattcgcg cggttctcaa ctccgacaaa gatattgata ttccgacaagc ctggctgcac 3840  
 ggaattgggg ggaagattgc ctatgagcat cccgagctgc gctgcacgtc cgtcgatctc 3900  
 gatgcgcaga gcaacgactg cgggcatctc gcgacgctga tgcgtgcgaa tatccgagag 3960  
 gatcaagttc cgcctccgca aggcacggta tgggcgcgc gccatgctc tcacaagatc 4020  
 ccatccgcac ccgatgtggc gttccgtgcc gacgcaacct attctgacac gggcgggctc 4080  
 ggccgagctg gactgcaggt ggcgggattg ctgcgcgcgc ccggagcgcg ccatctcgtt 4140  
 ctgctgggac gcagcgagcg tctctggcca caactggaa gtgtcgacgt caagatcatc 4200  
 catgcggagc tggcggaacc gcacgagcta tcggatgcgc tcgcgatcat cgcgcgcgac 4260  
 atgcccgcgt tgcggggcgt gttccatctg gcaggcagcg tggcgagcg catgctgctc 4320  
 aatctacga cgaacgcgtt cgaagccgcg attgctccga aagtgcggcg cgcgtggaaac 4380  
 ctgcaggaac tcacgcgcgg cgcgcgcgtg gatcaatttt ttctcttctc ttccgccagc 4440  
 gcgacagtcg gatctccggc ccagggcaac tacgcgcgcg gcaattcatt tctcgacgcg 4500  
 ctggctcacc tgcgcgcgcg ccagggtctt cccgcgcgta gcatcgctg gggacgcgtg 4560  
 accaggttgc gtttggcgcg acaggcgaa cgcggagacc gtctggcgcg cgcgcggcatc 4620  
 tccgttattc aaccgcaaca gggattgcgc gcgctctaca aagctctaca cgagattcgg 4680  
 ccgcagctcg ctgtcatgaa cttcgatata gcgcagtggc tccgttacta tccgtcggcc 4740  
 gcatcgatgt cctctgtggc cggcatcgca cccgcggcg cggaacacaa accgcgcgc 4800  
 gacatcgca cgcagctctc ggcaattcca gcggggcgcg agcgcgcgc gcgctggaa 4860  
 acgctgctga tgcacgaagc cggacacgtg ctgcgcttcg atccagcgaa actcgacggc 4920  
 agagcgagcg tgggtgatct cggattcgat tctgtgatgg ccctcgagtt tcgcaacctg 4980  
 ctggaagacg ggcctgcgct caagctttct gccaccctga tctggcgcta tccgacattc 5040  
 tccgcctcgg cgcagcatct cgcgcgaaag ctgcggctcg cgtgggaaag catggcgcgc 5100  
 aatgtgaac cttccacctg tgcctgcgct gctacccttg ctaccgcttg caccgcgcgc 5160

ggcgaaggacc ggagtcgccgc cgtcgcagac gatctcgacg ccgtcgcgaa ccagatcgcc 5220  
 gggttggggg acaaaagaaat cgaagctttg ttgaaacaga agttcgctca tttttcagga 5280  
 gctcccgagt ga 5292

<210> 118

<211> 6462

<212> ADN

<213> bacterie

<400> 118

gtgagttcga taticgagcg attccccaac cttacgccgt tgcagcaggc gtacctgacg 60  
 ctggagcaca tgcagcgacg tctcgatcg gcccgaacgcg acgcgcgcga acccatcgcg 120  
 atcgtggggtc tgggctgcgc gtttccgggc ggcatggggc ccgatgagtt ctggcagatg 180  
 ttgcgcagtg gagtcgatgc tattcgtgag gtaccgcctg gacgatggga cgaggagtcg 240  
 gtccggcgca tcttgaatc gtgaaacccc gccacgcgcg tgaagattca agccgattt 300  
 ctcgattcca tcgatgggtt cgacaacgat ttttccggca tttccgcacg cgaggcgctc 360  
 agcattgatc cgcagcagcg gctgctgttg gaagtggcgt ggcgagcact ggaggatcg 420  
 gggcagacga tggaaagggc ctccggcagc cgcacggcggt tcttcgtcgg gatccacagc 480  
 caaagcagcg actatttctg gatgcagacc gccgatggcg cgcgcacatga tccgtatacc 540  
 gccaccggca cggcgcatag cgtgatcgcc ggccgacttt cctatttgct gaacttgcaa 600  
 ggaccacga tcgcctcgca cacggcctgc tcttcttcgc tggcggcggt tcatctggcg 660  
 tgccagagcc tgcgcagcgg cgaagtgtacg ctggccgtgg ccggcgaggt gaactcgccg 720  
 ttctcgcggg agtttatgta cgccacctcg aagatgggaa ccgcctcgcc cagcggtcgc 780  
 tgccgcgcct tcgacgcggc ggccggacgcg atcgtgttcg gagaaggctg cgcgctgggtg 840  
 gtgctgaagc gctgttcgca tgcactcgcg gccggagacc ggggtgtggc cgtggtgcgc 900  
 ggctccgcgg tcaatcagga tggccgctcg gccgggctca ccgctcccaa tgcgtgtct 960  
 cagcaggtgc tcatccggtc ggcattggcc aatgcgggcg tcgcggcgca gcagatcggt 1020  
 tacatcgaag cccatggcac ggggactccg ctcgcgcatc ccacagatg cgaggcgctg 1080  
 gcggaacacg tcggcctccc gcgacctgtc ggcatgtgtg gcgcggtcgg gtccctgaaa 1140  
 tcgaacatcg gccacctgga gggagcgcca ggcatagcgg gattgattaa agcgggtgctc 1200  
 gcatttgatc acgagacgat accgccgagc ttacacgtga gacagctgaa cccgaatac 1260  
 cggttggagg gaacgtcgct cgacattgtg aaggaaagtc ggccgtggcc cgcgggttcg 1320  
 agacgaaggt ttgcggcgct cagcgctgtt ggttggtccg gcacgaacgc gcgatcgct 1380  
 cttagaagag cggcgccgac tggtagaggc gaagctgcga gcgggttcca ttcgccgacc 1440  
 ccgcgcgcgg ctcgcggggc ggctgtcccc ctccggagg gggacactgg gggcactccc 1500  
 gacattgcag gcaactccga cactgcagac actcccgaca ctcgagacac tcccagact 1560  
 gcagggactg caggcactgc ggcaactacg ggcatcgac acgcgatgta tgtgtctccg 1620  
 ctgtccgcgc atggtcgga cgaactgcgt cgggtggcgc gggcatacgg ggaattgctg 1680  
 acagcgtcgc acgcaccgag cctgcgtgat cttgtctaca cggccgcagt ccgccgcagc 1740  
 catcacgat cgcggctcgc tgtttccggc agaacggcgt aagaactggc ggcgcagctc 1800  
 cagggaatca cgatcccttc ccagcgcagc aagacgggat tctgttcttc tggacagctc 1860  
 tcgcaatgga tcggaatggg gcgcagctgg atggaccgcg aaccggttat tcgcgaggcg 1920  
 ttggaacgct gcgagccgc catgcggcct tatgtggact ggtcgctgaa agaagaactg 1980  
 gcgaagctcg acccgctcga ggtcattcag cctgcgctct tcgcgctgca ggtcgccatc 2040  
 gccgattgtt ggcgttctcg gggaaatcgag ccggatgcgc tcatcgggca cagcatggga 2100  
 gaggtcgcgc ccgctcatgt cgcgggtgcg ctgacgctgc aggatcgccg cgggatcatt 2160  
 tgcagccgca gccggctgtt gagccggatc agcggcctgg gcgggatggc gatggtgag 2220  
 ctgcgcctcg cggaaatgtga ggcgctgctg tcgacttaca cggaaacgat atcgcccgcg 2280  
 gtgtcgaaag gaccacaact caccgtcatc tccggtgaag tcgaagccct gcccgaggtc 2340

gtccgcgcgc tggagcggcg aggcgtgtct tgccggcccg tgaagtggga ctccgcgcgc 2400  
 catagcccg aagtggacc atgtgtgcga gaactcctgc agtcgctcga cgggattcaa 2460  
 ccgcggcccg cgaccatacc tttttactcc acggtgaccg gcgcgacgct ggagaccacc 2520  
 agcctcgaca gcacgtactg ggctcgcaat ctgcgatcgc cggttcttgt ctggcaggcg 2580  
 atccgcgcat ttgccgacag cgggcacgat gtctttctcg agatcagccc tcatcccatc 2640  
 ctgctgcccc ccactcggcg caatcggcgct ctggttcctg ctctgcgcgc cgaccaggac 2700  
 gaacgcggct caatgctcac gtctgtggcg gccctctatg aggcctggga cactgtcgca 2760  
 tggcggaccg tgtacccttc cggcaattgc gtgcgcctgc ccgggtatcc ctggcagcgt 2820  
 cgtcgtttct ggctcgacgc tccccccgc cgacacgcga tcacgttggg caatccgctg 2880  
 ttgggaaac cgtcggaagc ctgcgacgaa cccggcaatt tcttctggga gacggaactc 2940  
 agtctcgtct ccgtgccttg gctggcgacg catcgcgtgc agggcggaagt cgtcttgcgc 3000  
 gctactgcgt atctcgatat ggctctgggc ggaactctcg agaccttcgg tgaagtcccg 3060  
 tgcgtgtcgt agcatgtgac tttcacacag atgctcatgt tgccgcgcga cggcagcgt 3120  
 acgttgcgac tggccatcgc ggtcgataga cccgggatgg cgtcgtttcg gatttccagc 3180  
 cggcaggcat cgacatgggt cctgcgatgt tccggggaca ttctgcagac cactcgggat 3240  
 gcatcgaccg tccgcgggga tctgcgggag acggtgcagg cccgctgccc cacagtgggt 3300  
 ccggcggcg agctgtggcg tcagatggcg gagcacggcg tcgagtatgg tccgcttttc 3360  
 cgcgctcgc agcagatctg gagtgttcca ggtgaggcga tccggcgctc gcgtagctcg 3420  
 gaaacgcgtt ccactgcgcc ggcttctctc gatgcatgtc tcgagatcat cgcgcggcg 3480  
 tttgtcccc ccggtggaaac ctggctgccc gccggcatcg accggatgcg ctggctgcat 3540  
 cccgcacgtt ccgttggttg gacgcatgcy cggctggaag gacctatcgc cgactctgctg 3600  
 ctgctggagc gagagggaca actggtcgcc cgcacgagg gtctgcgctg cgacgctctg 3660  
 gatcgtcgg agcgcgatga catcgccggc ttggttgcacg aactgcgctg ggtcgtcag 3720  
 ccgcacgcgc ctgcagagcc gccggcgccg cgacggcgcc ggtcatggtt cattgtcggc 3780  
 gctgtggata ggcgcctcac cgcattggctg cgcgtctacgc gcaaccgctt gacgcagacc 3840  
 tcgctggaaa agctcgatga actccagccg ccgctcgagg aaatcgctt ttgtctcgag 3900  
 caccgaacct catcgagacc cattctgcat ctctccaga ccttggggcg caccgcttgg 3960  
 cgtcaagcac ccgcctatg gctggtcacg ccggcgccgc agccggtcga tggacagatc 4020  
 ctgcaagccg gtatcgctca ggcccttttc tggggtttgg gccggaccgt gcattacgaa 4080  
 catccgggac tgaactgcac gctgatcgat ctcgatcccg ccggcgccga agaggaaactc 4140  
 ctgcacgaac tgcctcgcaa caacggcgag aatcaaatcg cctttcgccg ccggcgccgt 4200  
 tacgtcgcgc gcttggtctg gcacgaagcg gatatgcaac ccgcatgtt caaggccggc 4260  
 gatcgccgtt tccggtcga gatcgatgcc ccggaggtcc tcgacggctg cgcttgcgc 4320  
 gccacatcg gccgcgccccc gcaagccggt gaagtggaga ttgaagtctg ccgcgcggcg 4380  
 ctgaacttcc tcgacgttct gctgccttc ggcttatgc ccgacgatgc gcccgcgcg 4440  
 attgcggcca gcccgctct ggccggcgga ttgctcgggc gtatcgtggc catggggaaa 4500  
 ggctcaccg actttcgcat cggagatgaa gtctgtggcc ttgcgccttg cagttctcgt 4560  
 cgcttcgta ccaaccccg ctctcgcgtt gccttgaagc cggccaacat tccgcgcgaa 4620  
 caggccgcgc cctgcctat ccgctttctc acgcgcgatt accgctctc gcgagcgcg 4680  
 cggtggcg ccggcggaag agtctgatt cacgctgcca ccggcggtgt gggatggcg 4740  
 gcaatccga tcgcacagcg tgccggcgcg gagatcttc ctactcgcg gactccggaa 4800  
 aaacgcagat atctgcctc gctgggcatc ggcgatgtt ccagttcgcg cgtatggct 4860  
 ttctgggacg acatccgcaa ttggacgaat caagaaggag tagacgtcgt cctgaattcg 4920  
 ctctcggcg atctcgtgga ggcgagcttc gatctgtgc gcgatcatgg accgttcttc 4980  
 gagatcgcca agcgcgat ta ctatgccggc cgcaagctgg ggcttctgcc gtctctgaag 5040  
 aaactctcgt acacgctggt cgatttgcct ggcatgtccc tgaagcgccc ggcattgacc 5100  
 cgggagctgc tgccaggagat ggtcgcaaaa ttcgaaatcg aaactctggc gccctggaa 5160  
 acgcgagtga cgaccatcac cgaatcgggt gaggcgcttc gcacctggc gcaggcgcg 5220  
 cacatcgcca aaactcgtcat ggcgatcgca gattgcgcca atgcgcccac cgcaccctca 5280  
 cgctcggcgt tcgatagcga gggaaacctac ttgattacc gcggaactgg cggctcggct 5340

```

cttaccgctcg cagcgtggat gatcggaacgc ggcgcccgcc ggctggtgct gctgagccgc 5400
cgcgcgcctt caccgcaggt ccagcaagcc atcgccgtca tggacgcaga tgtccgcgacg 5460
gtgcaggccg atgtttctca gcgcgatgaa ctgcagccgc tgatctcttc catcgatcga 5520
ttgcgcggcg tgattcatgc cgcagccgtt ctgcagcatg cgctgctact gaaccagacg 5580
gaagcgcatt tcgcgaacgt gatggccgcg aaaatcgacg gtgcctggaa cctgcacttg 5640
ctcacccgcg actgcccgct cgatcatctt gtgctcttct cctccgctgc aggactgctg 5700
ggcgcccgcc cccaggaaa ctacgcggcc gcgaacgcct ttcttgacgc gctggcctac 5760
taccggaagg accaaggcct gccggcgctt agcatcggtt ggggtgcgctg gtcggaggtc 5820
gggtggtgct cgcgcagga caatcgcgga tcgcggctgg ctttgccgcg catggaaaac 5880
ctgacgcgcg aacacggcct cgctattctg gaacagctgc tgaacagctc ggcttgccac 5940
gtgcgcgcga tgcccatcaa tgcgcgcag tggcgcgagt tctatcccaa ggcggcgag 6000
tctgactgt tcgagctttt gcatgacgac gcgcgcgagc aagccgatgc gccaaacgcg 6060
ttgcgcgcg ggctgcaatc ggccgagcct cagaccgcga ggacattgct cgaagaacat 6120
ctacagcaag agctggcgcg cgtgctgcgc atcgactctc aaactatcga tccccctgcg 6180
ccgctgaaag aactcggtcg gatttccctc atggccctgg agtttcgcaa ccgtctcgaa 6240
ctcacactgg gtctcacgct ccccgcgacc ctgatttggg gtcacccac gctggccggt 6300
cttgcgccgc acctggcgct gcaaatggga ctgcgctggt tcgaagcgca ggcgcgggct 6360
gctgcggaag gagacagcgc gcgatgaaa actgcactca gcgggttggg cgacatgtcg 6420
gaagaagcag ccgtggctgc gctccgagga gcaaggtcgt ga 6462

```

&lt;210&gt; 119

&lt;211&gt; 5088

&lt;212&gt; ADN

&lt;213&gt; bacterie

&lt;400&gt; 119

```

gtgaggga aaatgcgcc catgtcgtcg gtcaaacctc cgctattggc gcggaacatg 60
cggcaaaaca tgcaggctt cgacctggtt cagccgaac ccacgccat ctcggcgatg 120
cgctgtcggt ttccggcgcg gcgcgaagaat ccggacgcct tctggacgct gttgaagaac 180
gggtgtcagc gtgtcacgcg ggtgcccgca gaccgctgga actcggacca gtaactccc 240
tccgatcccg atgctccggg caaggcgtat gcgcgatatg ccgccttctc cgaacgcatt 300
gacggttctg atcggaatc cttcggcatc tccccccgcg aagctctgaa catggatccg 360
cagcagcgcg tgctgctgga agtgctgctg gaagcggcag aggaagcccg catctctccc 420
ggccctctcg cgggcagcgc gaccgcgcgc tttgcggcct cctgcgcccc ggaactcggg 480
ctgttttcag acgcccagcc tgcgcgcac ggagcttggg cgggttccgg cgtggcgcat 540
agcatgttgg ccaatcgcat ctctatctg ctgcacctgc cgggtccgag catggcggtc 600
gatacggcct gctctccgcg gctcgtcgcc gtcacatcgg ctctgcacaa cctgcggccg 660
cgcaaatgct atcgcgcatc cgcgcgcgga gtgaacttga tctgaactc ccaggcgatg 720
atcgctttgt cgaaggctcg catgttggcg ccgcacggag gctgcaagac gttcagcgcc 780
gcagccgacg gttatgtcgc cggcgaggcg tgcggcatcg tgctgtcgaa gcggctctcc 840
gatcgctggc cgcgtgcgga tgccatccgt cgactcatcc cggcgtcgcc aatcaatcag 900
gacggcagga gcaatggcat cagcgcgcgc aatctgcagg cgcagaaggc ggtcctgcga 960
gaggcggtgg ccaacgcgca catcgatcca tcccacgtat cggttgatcga ggcgatggc 1020
acgggcacgt cgtgggcgga tcttatcggt atcgaagccc tgacgtcggt ctacagcgcg 1080
cgggactctg cgcttctgtc gctgggttcc gtaaaagaca acatcgcgga tctggaggcg 1140
ggcgcggaat tcgcggcggt gatcaaaagc gtaactcgcc tgacgactcg caccattctc 1200
ccgcacctgc attttcgccg gctgaatccg aacatctcac tggagcggcg ccgggttctgc 1260
atcgccacgg aatcgtcgcc gtggacgtcg gaaggacggc cgcgtctggc cggcgtcagc 1320
tcgttcgggt ttggaggag caacgcgcac gtcactctcg aagaggcgcc tgcaactccc 1380

```

ttgcggaagc cggtcacacg cccgcagcct ctactctgt cggcgcgca cgcagaagcg 1440  
 ctccgggaac tggccggcca ctccggagag ttctgcagc cgcaccgcga tgcgttgctg 1500  
 tccgacgttt gcttcaccag tcaggttgagg cgcgacgat atagtcaccg ctctgggatc 1560  
 accgcgcgag atgcggcgaga ggctgtagcg gcattggccg cggcgccgcg cgcggaagta 1620  
 tcggtgcgcg ggccggcgccg aatcgctttt ctcttcaccg gccaggggcg gcagtagcgc 1680  
 ggcattggggc cagagcttta taaaacgcag cctgtttttc gcgacgcgct cgatcgttgc 1740  
 gcgcattggc tccgtcccca gctcgatgtt ccgctgaccg ttctcttggt cgagtcggtt 1800  
 tccgcgttgc acgagacggc gtataccca cggcaatgt ttgccctgga atgggctctg 1860  
 ctgcagttct ggctgtcgct cggcgctccg ccgactacg tgctgggcca cabtctcgcc 1920  
 gagtatgttg cggcgtgtgt ggccggcgcc tttagcgtgg aggagcgctt cgggctgggt 1980  
 accgccaggg ggccgctggt caatgcgctt ccccgcgga aagcggtcat cgttcacgcc 2040  
 aatccgagcc gcatcgccgc gctcgccgcc aagtgggcag tcgcgcgcat gaatgcgccc 2100  
 gaccgcaccg tgatctccgg caccgctgca gaaatcgcg aagcgcaaga tgacctgcat 2160  
 cgcgcggggc tggaaaacgc agagctgaac gtatcgcat cgttccattc gccgctggg 2220  
 gatccgattt tggacaagtt cgaagcgctt caggttgcca tcgctatca gccgctgagc 2280  
 atccctctgg tgcgaacgt cagcggagcc gtattgccga aaggcacagc actgcacgcc 2340  
 cgtcactggc ggcgacagtt gcgcgaaacc gtgcagtttg aaagcgcgat cgcgaacctg 2400  
 gcggaccgcg agtgcaagct gtttctggaa atcgcccgcc atcccacgct caccacgctg 2460  
 gggcgatatt gtctgcccga tgacggcgcg gcttggtcgc actccctatc taaggagcca 2520  
 tcggatttgg ccgtgtgctt ggaaagtctt gtcggcgctgt ttaccgcggg cgtgaatccc 2580  
 gactggcgcg gtctctatgc cggggaatca cccagccgcg tcgcgctgcc gacgtatccg 2640  
 tttcagcgtg acaccttcag cctgagacgc gtaccgcga gagagccggc gcgcggcgcc 2700  
 atggtgggag cgcgcctcaa cagcgcgttg ggcgatgtca tcttcgaaaa ttccgttaacc 2760  
 accgagacgc ctctgtccca tgagcacgtg atctacgacg cggctattgt gcccgcgccc 2820  
 tggcacgtgt cggcatttct cgaagcgcca caggaaagct tcgctcgccg tccctcgccc 2880  
 gtctccgatg tcatgatcg cagggcaact gccatcccg cggatacgcg ggtcacggtg 2940  
 caagcgattg tcacaccccg caggacggc gaagcaaaag tgcaggtctt cagccaggat 3000  
 ggcgattcgt ggaagctcca caccggaccc agtctgcgc cgcgcactgc cggcgccgtt 3060  
 catttcgagc tgcggcgca gccctccgaa gtcatctcg cgcattcggt ctacggcgcg 3120  
 atgaacgcac cggcgctcga tcttgccccc gccttcagtt ggggtgaaga agtctggcgt 3180  
 cgcgatggcg aggcgctggg gcgaatcgct ctgcgggtgg ctgaggaagg cgcgaacgct 3240  
 tacggcgctc accccggcct gatcgattct tgttttcaag tattcggagg gacttgcccc 3300  
 cgggagcggt gccagcccg cgcatacgt cccgctcgga tcgaagcggc gcgcttctac 3360  
 cgtccgcggc caggttctct gcgctgtcat gcgcgtctgc gcccgagctc gaggcgccc 3420  
 tctgtcgggt atctgacgct ggttgaagag accggcgcg tcatcgccga gttttccgga 3480  
 ctggctgtta tgcatggcgg tacgctgcaa tcgcacagt cgttgctgca ggaatgcaag 3540  
 tggcaggagt gcgagcgatc gacaacgttg aagtcgacg gccctggcaa gccggaggac 3600  
 tggttgctgt gtcccgcgcc agacgatgtc gccggttga tgccgcaaga gctgcgcgtc 3660  
 gtgtccggcg tcaactctcg ccaggcgctg gaacagacc agactttggt cggcgccccc 3720  
 gcgcggctct ggctgatcac gcgcggcggt catcgcatca gtgatgacga tgcacatccc 3780  
 gtcgactctt tccaggctcc actgtgggga ctccggcagg cgcacgcgc cgagcatccc 3840  
 gagctgtggg gcggcctgat gcacctcggt tgcgacaat ccgacatcgc ccccgccatg 3900  
 ctgctgggat aaatccgta tgccggcgac gacaagcgca tcgcattgcc caacggagc 3960  
 cgcacgcttc cgcgctggtt gcggcacaag gaaacgtgca agcggcgccg tggcatttca 4020  
 gccgacggcg tctatctgat caccggcggt ctccgcgcat taggacgaag ggtggcacgc 4080  
 cgcctgatcg agcaagcgcc gcgcgctgt gtactggctc gccggcatc ggaggcagtt 4140  
 gccgatctcg agcaactcgg ggctgcagtc atgtgttgcg cttgcgatgt gacttccgag 4200  
 caacagctgg cggcgtgctt ggcggaaccc cgcacccagc cgtcgctgtg agtcgtgatc 4260  
 gccgacggcg tctcgatga cggggtagtt acagaacaga cgtggcgctg tttcggaag 4320  
 gtgtcgccgc cgaagctgca ggggtccctg aatcttcacc agctcactcg ccaccatgcg 4380

```

ctcgactttt tcgtactctt ctcttccgcc gcttcgctgc tcggttccgc cggacagagc 4440
aattactcgg cggccaacgc atttctcgac agccttgccc acatcgccgc cgcgcaagga 4500
ctaccggcgc tgagcatcaa ttggggacca tggcggggag aaggcatggc cgcgcgcctc 4560
gcgcgccgaa gcctgcccgg ggtaccgctg ctgcccgcgg aagtgggtgc gcgcatcttc 4620
ggcgatctgc tggcgagac tgccgctcag atccgggtgt tccaagtctc cgccgaaaaa 4680
aggcggagcc cggcgagcga tcccggtctc atccagcaac tcaccgaagc tgcgccggag 4740
cggcgccagg aactgtgcga gatgcgcctc cgcaagcagg ccggcgccgt gctggcgctc 4800
gatgcgtcca agacgctcga cccgcgcggg ccgctcaagg aatacggact cgattcgctg 4860
atggcgctgg atctggcgcg cgccatcggg gagctgggtg gcaagagcct tcccgcgaca 4920
ttgctatacg accatccgac cgtcgagaaa ttggccggcc atgtcctccg cgaactcgga 4980
ctcgacgtcc ccagcgatcc cctcgtcgat gaagtgcggc agctgtccga gcaggagatg 5040
gcggcggtca tcacggaaac cttgcaccat ctgggagagg aacgatga 5088

```

&lt;210&gt; 120

&lt;211&gt; 4306

&lt;212&gt; ADN

&lt;213&gt; bacterie

&lt;400&gt; 120

```

atgagcgatc tcactcctct tcaacaggcg gtctggcgcc tcaagcgcac gcgagcgct 60
ctcgacgaac tggagagcgt ccacaacgaa cccatcgaga tcgttggcat ggtctgcgcg 120
tttccggcgc cggactcgcc ggaagcattt tggcagctcc tgcacgatgg catcgatgcc 180
atcccgcaaa ttccctcggg ccgttgggat gccgatgctg tttacgatcc cgatcccaac 240
gcgcggggaa agatgtacac gcgtctgggg ggattctctg atbggtccgt cgacggcttc 300
gacgcggggt tcttcggaat caocccgcgc gaggtcgccg gtctggatcc cgacgcgcg 360
ctcgtcgtcg aggtggcatg ggaagctttg gagcgtgcgg gtccggccgc cgacagtctc 420
gcgggacgag acaccggagt gttcatcggg atcagcaccg acgactacag ccggctgaaa 480
cctaccgatc cggcgctcat tgacgcctat accggtaccg gaaccgcgtt cagcactgcc 540
gcgggacgga tctcctatct gctggggttg cagggacgga acttcccgtt cgacacggcg 600
tgctctctct caactcgtgc ggttcactct gcgtgccgca gcttgcaagt gcgagagtgc 660
agcatggcgc tggccggcgg cgtgaacctg attctggcgc cggaaagcac gatctacttc 720
tgccgcctgc gggccatggc ggcgatggc cgttgcaaaa gtttcgctgc ctccgccgac 780
ggttacggcg cggcgaggag atgcggaatg ctggtgctga agcggtgttc cgatgcgacg 840
cgtgacggcg atcgatatct ggcgctgatt cgcggatcgg ccgtcaacca cggcgccgcg 900
agcaacggcc tcacggcgcc gaacggtccg gcgcaggaag ccgtgatcgg ggcggcgctc 960
aagaacggcg gcatggcccc gcgcgatgtc gattacgtgt gattacgtgt aaccgggacg 1020
ccgtgggag atcccatcga actcggggcg atggcagcgg tgctggggca gggcggtgcc 1080
ctcgattctc cgttgatcgt cgggtcgggt aaacacaaat tcggccacct ggaggcggcg 1140
gcaggatcgc cggcgctgat caagaccatt ctgcacctgc agcacgaga gattccgcct 1200
catctgcatt tcaacgcgcc caaccgcac gtactctgga atgagctgcc gctaaagata 1260
gccacgcgat gttcgcatg gccctccaa cgcgcgcccc gaggtgccgc ggtgagctcg 1320
ttcggaatca gtggcaccaa ttcgcacgtc gtctcgcag aagcgaagac gaatgtagaa 1380
gcgaagacga atgtagaggc gaagacgaat gttagaggca agacgagtga agaggtcaag 1440
gcgagtgtag aggcacaaagg gaatgtggag gctaaggcta gtctagtgt cccctctctc 1500
gagggggaca gccgcccgcg aagcggcgcg ggggggtcgg gccggccgcc cagccgcgag 1560
gaagtgcggc tcccgatca actccatgcc gaagacggcg ccgaataacct cctaccgctt 1620
tcggcgccgc ctcgcagggc tctgcgcgat ctgcggcgcg cctatcgaga tctggcgctt 1680
cacgctccgc tctccgcgt gtgttcggcc gccagcctga cgcgcagtca ctacgaacat 1740
cgcgacgctg ttgtggccctc atccctgccc gaggtaactc aattgctcga ggccttccgg 1800

```

cgcaatgaaa ccaatcgccg cgtcgccacc ggtttcgccg atcccgaggt tcgtccgaaa 1860  
 ctgcgcttca tcttttccgg ccaggcgga cagtaccgcg gcatggcgta tgcctgttat 1920  
 tccgacgagc ctgtcttccg atcgccgcat gaacgttgcg acgcccgcct ccgcagcttc 1980  
 ttggaatggc ggcttgcgga cctgctcgcc gacgagtcgg gagcatggct gagccagatc 2040  
 gatcgcgctg agcctcgctt gttcgccgtt caaatcgccg tggtcgaact gctgcaatcc 2100  
 tggggaaatc gcccgagcgg cgtggccgga cacagcatgg gagaagtggc ggcggcccat 2160  
 gtgcagacca ttctaccct ggaggacgcg gcccgcatca tctgtcgccg cagccggctg 2220  
 ttgctcgacc ttcgccgccc gggagcgatg gctctggtcg aactgcgcgt cgatcgggcg 2280  
 aaggccgtgc tgcgtgaacg cggctcact actgtttctg tccgcccagc caacggacca 2340  
 cgcgacacgg tgttctcggg agaccgtgtg gctctcgagc atttgaagga cgacttcgag 2400  
 aggcgcggcg ttttctcgcc gctgattcag gtggaatgct cttcacacag ctgcaggggt 2460  
 gaccgcgtcg agaagcaatt gcgcaggaa ctcggccgcg ttattgcaaa acgttccgcc 2520  
 gtgcggttct tctccacggt tgaaggacag ttgagcacgg gcgagggcgt cgacgcgtcg 2580  
 tactgggtag ccaatctgcg acagccagtc cgtttctggg agtcggttga ggcgatggct 2640  
 ggtgatgagt tcacgcagtt cctggagatc agtcgcgcat cgtgtcgatg gccgtcgatc 2700  
 gaggatagtc tgcggacgct cggcataaac ggactggttc gcccgctact gcgcgcgac 2760  
 gaaccggagc ggcttgagct gctcgagttg ctcgccgcgc tctacgtgaa tgggcagcgt 2820  
 ccggactggc gcgcgtcgcc ttgcgtctcc gacacgcgcg tggatctgac gacgatcccc 2880  
 tggcagcgcg agcgtctctg gttcgcgacc tgcacgcggc gaagtgttgc ggcagttggc 2940  
 ggtcatccgc tgcctcggtc caaggtcgag attgcgctgg gcgcggacac acacgtctgg 3000  
 gagtcctgtc tctctctgga tgcgctgcgc tttctcgccg atcaccggct caacgagctt 3060  
 tgggtgcttc ccggtgcgcg ttatgtggag atggcgctgg ccgcagccaa ggaagtgttc 3120  
 cggggtgctg cgcgctcgga agagatccgg ttgaaacaaa tgcgtggtgt tctctccgcg 3180  
 ggcgcctcgc gagtgcaggt cactactcag ggacacgcat tccgcattct cagttcgcc 3240  
 gaaggcggtt ccgattggac cgagcacgcg cgcggcacca tggctgcggc gccggacaag 3300  
 gtcgcgcaca cggtagacct gccacacct ggggatcgca tccgagggca tgacttctat 3360  
 cgcgccttgc catcgacggg gatgcattac ggcgacacct tccgcggcat cgcggaagtg 3420  
 tggcgcgcg cgcgcgaggg agtggcgcca ctgagcgctg cggatgccgt tccggaagca 3480  
 gagtccgggt acacgcttca tctctgcctg ctcgatgctt gtttgcaggt gctgggcgcg 3540  
 acgttgggag gcgaaggcag gcgcggctct tgcgtgcctg tcccatcga acggttgcac 3600  
 tgtttcggca gaccgcggcg cgatcttagg gtgcgacgcg ggctgacggg gcggtcgag 3660  
 ggcgatgtca ccctgtgtga tgcggaaagg cactcatcc tcgaggtcca aggcctgcgt 3720  
 gccaggaac tggagcgcca atccgaatgg ttccagctca tggaaatggga gccgcagctg 3780  
 ctggcgagga gtccaacggc aacggtgtcg ggtgcattgc tggctattgc cgatcgccg 3840  
 ggcacgcag ccgcggtggc gcgagggctg ggcacaaaca cggttgtgat ttcgggtcgc 3900  
 gatcgcgaga taccgataca gccctaccgg ggogtcattc actcggggag cctggatgag 3960  
 accgaggatg agaccgatcc ttcgggtgcg gggggaaccg cctgcgaagc catatttgcg 4020  
 atcgttcaag aattccgaggt gggacgcata cagctgacga aacaagcgtc cgacgcgaa 4080  
 tccgacatc ccggaatctg gctgattacg gcggcggttc atcgcgagca tctgcagatg 4140  
 ccggtgtgtc ccgcgcgggc accggtgtgg ggtctgggac gtaccatcgc ggcgcagcat 4200  
 cccgagttcg cttgcacctg catcgattct gacactgcgc gtgaagtcca ggtgcaggcg 4260  
 ctctgcgag agattctcgc ggggagttct gaacgtcagg gcccg 4306

&lt;210&gt; 121

&lt;211&gt; 1537

&lt;212&gt; PRT

&lt;213&gt; bacterie

&lt;400&gt; 121



Leu Gln Cys Pro Glu Ser Ala Val Asp Leu Gln Gln Pro Leu Val Arg  
 1 5 10 15  
 Met Gly Leu Asp Ser Leu Met Ala Val Gln Leu Arg Asn Arg Ile Asp  
 20 25 30  
 Thr Asp Leu Arg Val Leu Leu Pro Met Val Arg Phe Leu Asp Gly Pro  
 35 40 45  
 Ser Val Ala Glu Leu Ala Arg Asp Leu Ser Asp Leu Ser Gly Leu Ser  
 50 55 60  
 Glu Arg Thr Thr Val Ala Pro Glu Pro Ala Ala Gln Ala Ser Val Pro  
 65 70 75 80  
 Ala Leu Ser Tyr Pro Leu Ser Ala Gly Gln Gln Ala Leu Trp Phe Ile  
 85 90 95  
 Tyr Arg Ser Ala Pro Glu Ser Pro Ala Tyr Asn Ile Ala Trp Ile Ala  
 100 105 110  
 Arg Ala Arg Gly Ala Phe Asp Pro Gln Ala Leu Arg Arg Ser Leu Gln  
 115 120 125  
 Asp Leu Val Asp Arg His Pro Ala Leu Arg Thr Thr Ile Ala Glu Ser  
 130 135 140  
 Gly Gly Ala Pro Val Gln Thr Val His Ser Ser Val Pro Val Asp Phe  
 145 150 155 160  
 Glu Val Ile Pro Cys Ser Pro Asp Asp Glu Ala Val Leu Ile Asp Gly  
 165 170 175  
 Val Phe His Ala Pro Phe Asn Leu Gly Glu Asn Cys Phe Arg Ser Arg  
 180 185 190  
 Leu Leu Val Gln Ser Gly Lys Asp Gln Val Leu Ala Ile Val Val His  
 195 200 205  
 His Ile Leu Ala Asp Phe Trp Ser Leu Leu Val Met Val Asp Glu Leu  
 210 215 220  
 Arg Ser Ile Tyr Leu Ala Arg Thr Ala Gly Gly Pro Pro Val Ala Pro  
 225 230 235 240  
 Pro Val Ala Ser Phe Ala Ala Phe Val Arg Trp Gln Asn Glu Leu Leu  
 245 250 255  
 Ala Gly Thr Glu Gly Glu Arg Leu Trp Asn Tyr Trp Ser Ser Gln Leu  
 260 265 270

Ser Gly Gln Leu Pro Val Leu Asn Leu Pro Ser Asp Arg Pro Ser Pro  
 275 280 285

Pro Val Gln Ser Phe Arg Gly Asn Ser His Ser Phe Arg Ile Glu Pro  
 290 295 300

Ala Leu Thr Ala Lys Leu Lys Ala Leu Ala Arg Arg Gln Asn Ala Thr  
 305 310 315 320

Leu His Ala Thr Leu Met Ala Ala Phe Gln Val Leu Leu Ser Arg Trp  
 325 330 335

Thr Ser Gln Glu Glu Ile Leu Thr Gly Thr Leu Thr Asn Gly Arg Thr  
 340 345 350

Gln Pro Glu Phe Ala Asp Leu Val Gly Tyr Phe Val Asn Pro Val Ile  
 355 360

Leu Arg Gly Glu Leu Ser Gly Asp Pro Asp Phe Asn Thr Val Leu Ala  
 370 375 380

Arg Ile Arg Gln Thr Leu Leu Gly Ala Ile Glu His Gln Glu Tyr Pro  
 385 390 395 400

Tyr Ala Arg Ile Val Glu Arg Leu Gly Pro Gly Leu Arg Val Leu Phe  
 405 410 415

Val Leu Gln Gln Pro His Arg Ile Pro Glu Ser Val Pro Phe Met Leu  
 420 425 430

Gly Gln Ser Gly Gly Arg Met Ala Trp Gly Ser Leu Thr Leu Glu Ser  
 435 440 445

Leu Ala Met Pro Leu Arg Gln Ser Arg Phe Asp Leu Asp Leu Met Met  
 450 455 460

Val Glu Thr Asp Gly Gly Leu Ser Ala Phe Leu Gln Tyr Asn Thr Asp  
 465 470 475 480

Ile Phe Asp Ala Ala Thr Ile Glu Arg Leu Ser Leu His Phe Ala Val  
 485 490 495

Leu Leu Glu Gly Ile Ala Glu Asn Pro Ala Cys Pro Val Val Asp Leu  
 500 505 510

Pro Leu Leu Thr Thr Arg Glu Arg Ile Gln Leu Leu Glu Glu Trp Asn  
 515 520 525

Ala Thr Ala Ala Glu Phe Pro Ser Gln Cys Val His Glu Leu Phe Glu

| 530                     | 535                 | 540                     |
|-------------------------|---------------------|-------------------------|
| Ala Gln Val Glu Leu Thr | Pro Asp Ala Ile     | Ala Leu Ser Phe Gly Glu |
| 545                     | 550                 | 555 560                 |
| Gln Asn Leu Thr Tyr Arg | Glu Leu Asn Gly     | Ser Ala Asn Arg Ile Ala |
| 565                     | 570                 | 575                     |
| His Tyr Leu Arg Ser Arg | Gly Ala Gly Pro Gly | Glu Met Val Gly Ile     |
| 580                     | 585                 | 590                     |
| His Val Thr Arg Ser Leu | Glu Thr Val Ala Gly | Leu Leu Gly Val Leu     |
| 595                     | 600                 | 605                     |
| Lys Ala Gly Ala Ala Tyr | Val Pro Leu Glu     | Pro Glu Tyr Pro Ala Gln |
| 610                     | 615                 | 620                     |
| Arg Leu Arg Leu Met Leu | Glu Glu Thr Arg     | Pro Val Val Val Leu Asn |
| 625                     | 630                 | 635 640                 |
| Val Thr Glu Ser Glu Val | Trp Thr Gln Pro Asp | Thr Asn Pro Asn Pro     |
| 645                     | 650                 | 655                     |
| Leu Ala Thr Pro Ala Asp | Leu Ala Tyr Val Leu | Tyr Thr Ser Gly Ser     |
| 660                     | 665                 | 670                     |
| Thr Gly Arg Pro Lys Gly | Val Gln Ile Thr His | Gln Ala Val Val Asn     |
| 675                     | 680                 | 685                     |
| Phe Leu Ser Ser Met Arg | His Glu Pro Gly Ile | Ser Asp Arg Asp Thr     |
| 690                     | 695                 | 700                     |
| Leu Leu Ala Leu Thr Thr | Phe Met Phe Asp Ile | Ser Ala Leu Glu Ile     |
| 705                     | 710                 | 715 720                 |
| Phe Leu Pro Leu Ser Ala | Gly Ala Arg Val Val | Val Ala Asn Gln Glu     |
| 725                     | 730                 | 735                     |
| Thr Ala Val Asp Gly Glu | Arg Leu Ala Arg Glu | Leu Ala Arg Ser Lys     |
| 740                     | 745                 | 750                     |
| Ala Thr Met Met Gln Ala | Thr Pro Ala Thr Trp | Arg Leu Leu Leu Ala     |
| 755                     | 760                 | 765                     |
| Ser Gly Trp Pro Gly Asp | Arg Arg Leu Thr Ala | Leu Cys Gly Gly Glu     |
| 770                     | 775                 | 780                     |
| Ala Leu Pro Arg Asp Leu | Ala Asp Arg Leu Leu | Gln Arg Thr Ala Ala     |
| 785                     | 790                 | 795 800                 |

Leu Trp Asn Leu Tyr Gly Pro Thr Glu Thr Thr Ile Trp Ser Ala Ile  
 805 810 815

Gln Arg Val Thr Thr Gly Asp Gly Pro Val Ser Ile Gly Arg Pro Ile  
 820 825 830

Ala Asn Thr Gln Leu Tyr Val Leu Asp Asp Arg Met Gln Pro Ala Pro  
 835 840 845

Ile Gly Val Ala Gly Glu Leu Tyr Ile Gly Gly Ala Gly Leu Ala Arg  
 850 855 860

Gly Tyr Leu Asn Arg Pro Glu Leu Ser Ala Asp Lys Phe Val Ala Asn  
 865 870 875 880

Ser Phe Asp Pro His Gly Thr Arg Leu Tyr Arg Thr Gly Asp Leu Ala  
 885 890 895

Arg Arg Gln Arg Asp Gly Ala Leu Glu Tyr Leu Gly Arg Ile Asp His  
 900 905 910

Gln Val Lys Ile Arg Gly Phe Arg Ile Glu Thr Gly Glu Ile Glu Ala  
 915 920 925

Ala Val Arg Ser His Pro Ala Val Arg His Ala Val Val Thr Ala Arg  
 930 935 940

Glu Asn Asp Ala Ala Gly Lys Tyr Leu Ala Ala Tyr Ile Val Pro Leu  
 945 950 955 960

Ala Asp Gly His Arg Ala Thr Ala Ala Ala Asp Thr Phe His Asp Arg  
 965 970 975

Val Glu Ser Glu His Val Thr Gln Trp Gln Ser Val Trp Asp Thr Thr  
 980 985 990

Tyr Glu Gln Asn Ala Pro Asn Ala Asp Pro Glu Phe Asn Ile Val Gly  
 995 1000 1005

Trp Arg Ser Ser Val Thr Gly Glu Pro Ile Pro Ala Ala Glu Met Arg  
 1010 1015 1020

Glu Trp Val Gln Asp Ser Val Asp Arg Ile Leu Ala Ser Arg Pro Arg  
 1025 1030 1035 1040

Arg Val Leu Glu Ile Gly Cys Gly Thr Gly Leu Leu Leu Phe Arg Val  
 1045 1050 1055

Ala Pro His Cys Ser Glu Tyr Trp Ala Thr Asp Phe Ser Gln Lys Ala  
 1060 1065 1070

Leu Asp Tyr Ile Ala Ala His Ala Asp Arg Thr Gly Leu Ala Asn Val  
 1075 1080 1085

Arg Thr Phe Arg Gln Ala Ala Asp Asp Ala Cys Glu Ile Asp Ser Arg  
 1090 1095 1100

Ser Cys Asp Ala Val Val Leu Asn Ser Val Ile Gln Tyr Phe Pro Gly  
 1105 1110 1115 1120

Glu Ala Tyr Leu Arg Arg Val Leu Ala Glu Ala Val Arg Val Val Lys  
 1125 1130 1135

Pro Gly Gly Ile Val Phe Val Gly Asp Val Arg Ser Leu Pro Leu Leu  
 1140 1145 1150

Glu Thr Phe Tyr Ala Ser Leu Glu Val Gln Arg Ala Pro Ala Ser Leu  
 1155 1160 1165

Thr Arg Asn Glu Phe Arg Gln Arg Val Arg Ser Leu Ala Ser Gln Glu  
 1170 1175 1180

Glu Glu Leu Val Val Asp Pro Ala Phe Phe Phe Ala Leu Arg Glu Gln  
 1185 1190 1195 1200

Ile Pro Glu Ile Gly Arg Ile Glu Ile Leu Pro Arg Arg Gly Arg Ser  
 1205 1210 1215

His Asn Glu Leu Thr Arg Phe Arg Tyr Gln Ala Ile Leu His Ile Gly  
 1220 1225 1230

Ser Arg Glu Ala Glu Glu Pro Glu Ser Asp Arg Arg Arg Cys Gln Thr  
 1235 1240 1245

Ala Ala Glu Ile Arg Arg Val Leu Thr Asp Ala Gln Pro Glu Leu Ala  
 1250 1255 1260

Ala Phe Thr Glu Ile Pro Asn Ala Arg Leu Thr Ala Glu Ser Ala Ile  
 1265 1270 1275 1280

Val Thr Trp Met Asn Gly Asp Glu Ala Pro Glu Thr Leu Gly Glu Leu  
 1285 1290 1295

Arg Asp Arg Leu Arg Gln Thr Ser Pro Ser Gly Val Asp Pro Ala Asp  
 1300 1305 1310

Leu Trp Arg Met Asp Glu Asp Leu Pro Tyr Arg Val Ala Ile Asp Trp  
 1315 1320 1325

Ser Ser His Gly Pro His Gly Arg Phe Asp Ala Thr Phe Cys Arg Ala

1330                      1335                      1340  
 Ala Ala Gly Pro Pro Ala Ser Arg Pro Arg Arg Arg Leu Ala Gly Pro  
 1345                      1350                      1355                      1360  
 Tyr Thr Asn Asp Pro Leu Arg Ala Val Tyr Thr Arg Thr Val Val Pro  
                     1365                      1370                      1375  
 Gln Leu Arg Thr His Leu Lys Glu Lys Leu Pro Asp Tyr Met Ile Pro  
                     1380                      1385                      1390  
 Thr Ala Trp Val Val Leu His Glu Met Pro Leu Thr Pro Asn Gly Lys  
                     1395                      1400                      1405  
 Ile Asp Arg Asn Ala Leu Pro Asp Pro Glu Pro Ser Arg Arg Ala His  
                     1410                      1415                      1420  
 Ala Glu Ala Phe Thr Pro Pro Glu Thr Pro Val Glu Gln Val Leu Ala  
 1425                      1430                      1435                      1440  
 His Ile Trp Gly Glu Val Leu Gly Met Asp Gly Ile Gly Val His Asp  
                     1445                      1450                      1455  
 His Phe Phe Asp Ser Gly Gly His Ser Leu Leu Val Thr Gln Met Ile  
                     1460                      1465                      1470  
 Ala Arg Val Arg Asp Met Leu His Val Glu Val Pro Phe Arg Thr Val  
                     1475                      1480                      1485  
 Phe Asn Ala Pro Thr Val Arg Gly Phe Ala Val Ala Ile Gln Asp Gly  
                     1490                      1495                      1500  
 Val Asp Pro Gly Trp Ala Arg Arg Ala Ala Asp Leu Leu Ile Ala Val  
 1505                      1510                      1515                      1520  
 Ser Gln Met Ser Asp Val Gln Ile Glu Arg Met Met Ser Ala Ala Gln  
                     1525                      1530                      1535

Asp

<210> 122

<211> 2766

<212> PRT

<213> bacterie

<400> 122

Met Gln Asn Ser Ser Pro Asn Thr Ile Asp Leu Ser Leu Ala Arg Arg

|                                                                 |     |     |     |
|-----------------------------------------------------------------|-----|-----|-----|
| 1                                                               | 5   | 10  | 15  |
| Gln Leu Leu Asp Arg Leu Leu Gln Glu Asn Ser Pro Glu His Arg Ile | 20  | 25  | 30  |
| Pro Arg Arg Glu Asn Arg Asp Ala Ala Pro Leu Ser Leu Ala Gln Gln | 35  | 40  | 45  |
| Arg Leu Trp Phe Leu His Gln Leu Asp Pro Asp Ser Pro Ala Tyr Asn | 50  | 55  | 60  |
| Ile Pro Ile Ala Leu His Ile Arg Gly Pro Leu Asp Ile Arg Val Leu | 65  | 70  | 75  |
| Leu Arg Ser Leu Glu Ala Val Val Gln Arg His Glu Ser Leu Arg Ser | 85  | 90  | 95  |
| Cys Ile Gly Gly Val Asp Gly Glu Ala Arg Gln Ser Leu Leu Ala Arg | 100 | 105 | 110 |
| Val Thr Leu Glu Leu Pro Val Val Gln Ala Asp Gly Ile Ala Glu Ala | 115 | 120 | 125 |
| Arg Gln Met Ala Leu Arg Asp Ala Gln Ile Pro Phe Asp Leu Arg Lys | 130 | 135 | 140 |
| Pro Pro Leu Leu Arg Thr Lys Leu Ile Cys Leu Asp Asp Lys Gln Gln | 145 | 150 | 155 |
| Ile Leu Leu Leu Thr Leu Ser His Ile Ile Ala Asp Ala Trp Ser Val | 165 | 170 | 175 |
| Glu Thr Phe Val Arg Asp Leu Thr Arg Ser Tyr Glu Ala Phe Val Gln | 180 | 185 | 190 |
| Gly Arg Pro Ser Pro Leu Met Glu Leu Pro Ile Gln Tyr Gly Asp Trp | 195 | 200 | 205 |
| Ala Val His Gln Gln Thr Ser Leu Asn Gln Thr Ala Gln Gln Tyr Trp | 210 | 215 | 220 |
| Lys Lys Gln Leu Ser Gly Thr Leu Pro Phe Leu Asp Leu Pro Thr Asp | 225 | 230 | 235 |
| Arg Pro Arg Pro Ala Gln Gln Thr Trp Arg Gly Ala Val Glu Thr Thr | 245 | 250 | 255 |
| Ala Leu Gly Arg Asp Leu Thr Asp Gly Leu His Ala Phe Ala Leu Arg | 260 | 265 | 270 |

Glu Gly Ala Thr Val Phe Met Thr Ala Ile Ala Ala Phe Gln Val Leu  
 275 280 285  
 Leu His Arg Tyr Thr Ala Gln Glu Asp Ile Leu Ile Gly Val Pro Val  
 290 295 300  
 Ala Gly Arg Thr Gln Arg Glu Thr Glu Gly Leu Val Gly Cys Phe Ala  
 305 310 315 320  
 Asn Met Ile Val Leu Arg Gly Asp Leu Arg Asp Asp Pro Ser Phe Arg  
 325 330 335  
 Ser Leu Leu Ala Arg Thr Arg Asp Thr Ala Leu Ser Ala Leu Ser His  
 340 345 350  
 Gln Asp Phe Pro Phe Glu Arg Leu Val Glu Glu Leu His Pro Pro Arg  
 355 360 365  
 Asp Leu Ser Arg Ser Pro Val Phe Gln Val Ser Phe Ala Leu Leu Pro  
 370 375 380  
 Asp Ala Pro Ala Ile Thr Val Met Pro Gly Leu Thr Ile Ser Arg Glu  
 385 390 395 400  
 Tyr Met His Asn Gly Gly Ser Lys Leu Asp Leu Gly Val Thr Leu Glu  
 405 410 415  
 Pro Ser Gly Asp Gly Leu Met Ala Ser Ala Glu Tyr Asn Thr Asp Leu  
 420 425 430  
 Phe Asp Ala Ala Thr Ile Ala Ser Leu Leu Asp Ala Tyr Arg Thr Leu  
 435 440 445  
 Leu Ala Ser Val Val Thr Asp Pro Asp Val Arg Ile Ser Thr Ala Ala  
 450 455 460  
 Leu Leu Ser Pro Ala Val Arg Ser Arg Met Leu Glu Gln His Asn Ala  
 465 470 475 480  
 Thr Arg Arg Asp Ala Gly Pro Asn Gly Cys Ala His Glu Leu Val Glu  
 485 490 495  
 Ala Gln Ala Glu Arg Thr Pro His Ala Val Ala Val Val Phe Glu Asp  
 500 505 510  
 His Gln Leu Thr Tyr Ala Glu Leu Asn Ala Arg Ala Asn Arg Leu Ala  
 515 520 525  
 His Arg Leu Ser Ala Ser Gly Ala Gly Pro Gly Lys Ile Ile Ala Leu  
 530 535 540



Ala Met Glu Arg Ser Leu Glu Met Val Ile Ala Leu Leu Ala Ile Leu  
 545 550 555 560  
 Lys Ser Gly Ser Ala Tyr Leu Pro Leu Asp Pro Ala His Pro Lys Asp  
 565 570 575  
 Arg Leu Ala Arg Ile Leu Asp Glu Val Gln Pro His Ala Val Leu Thr  
 580 585 590  
 Gln Glu Ala Val Ala Glu Met Met Ala Met Met Ala Met Met Ala Val  
 595 600 605  
 Ala Val Glu Pro Glu Ala Ala Asn Leu Val Ser Gly Ser Lys Pro Asp  
 610 615 620  
 Asp Leu Ala Tyr Ile Ile Tyr Thr Ser Gly Ser Thr Gly Arg Pro Lys  
 625 630 635 640  
 Gly Val Glu Ile Arg His Ser Ser Leu Val Asn Leu Leu Arg Ser Met  
 645 650 655  
 Gln Arg Glu Pro Gly Leu Thr Ala Ala Asp Gly Leu Val Ala Val Thr  
 660 665 670  
 Thr Val Ser Phe Asp Ile Ala Gly Leu Glu Ile Trp Leu Pro Leu Ile  
 675 680 685  
 Thr Gly Ala Arg Val Ile Val Ala Thr Arg Glu Ile Val Val Asp Gly  
 690 695 700  
 Glu Arg Leu Thr Thr Leu Leu Asp Lys Ser Gly Ala Thr Val Met Gln  
 705 710 715 720  
 Ala Thr Pro Ser Gly Trp Arg Gln Leu Leu Asp Ser Gly Trp Lys Pro  
 725 730 735  
 Gly Lys Gly Phe Arg Val Phe Cys Gly Gly Glu Ala Leu Pro Glu  
 740 745 750  
 Leu Ala Arg Arg Ile Leu Asp Ser Gly Val Glu Leu Trp Asn Leu Tyr  
 755 760 765  
 Gly Pro Thr Glu Thr Thr Ile Trp Ser Ala Val His Lys Thr Gln Arg  
 770 775 780  
 Leu Gly Ala Ser Asp Ser Ile Val Pro Ile Gly His Pro Ile Asp Asn  
 785 790 795 800  
 Thr Gln Leu Tyr Ile Leu Asp Ser Arg Met Glu Pro Val Pro Pro Gly

|                                                                 |      |  |      |  |      |
|-----------------------------------------------------------------|------|--|------|--|------|
|                                                                 | 805  |  | 810  |  | 815  |
| Val Pro Gly Glu Leu Tyr Ile Gly Gly Ala Gly Leu Ala Arg Gly Tyr | 820  |  | 825  |  | 830  |
| His Arg Asn Pro Glu Leu Thr Arg Glu Lys Phe Arg Glu Trp Arg Asp | 835  |  | 840  |  | 845  |
| Arg Gly Arg Ile Tyr Ser Thr Gly Asp Leu Ala Arg Tyr Arg Ser Asp | 850  |  | 855  |  | 860  |
| Gly Ala Val Glu Cys Leu Gly Arg Val Asp Arg Gln Ile Lys Leu Arg | 865  |  | 870  |  | 875  |
| Gly Phe Arg Ile Glu Pro Ala Glu Ile Glu Ala Ala Ile Glu Thr His | 885  |  | 890  |  | 895  |
| Ile Ala Val Lys Gln Ala Ile Thr Val Val Lys Asp Asp Arg Leu Ile | 900  |  | 905  |  | 910  |
| Ala Tyr Leu Val Pro Ala Thr Gly Asp Val Arg Asp Leu Gln Ser Asp | 915  |  | 920  |  | 925  |
| Leu Arg Ser Trp Leu Ala Thr Arg Leu Pro Asp Tyr Met Ile Pro Ser | 930  |  | 935  |  | 940  |
| Ala Phe Val Ser Leu Ser Ser Leu Pro Leu Thr Pro Asn Gly Lys Ile | 945  |  | 950  |  | 955  |
| Asp Ala Asn Ala Leu Pro Gly Leu Pro Thr Thr Pro Val Ala Ala Arg | 965  |  | 970  |  | 975  |
| Glu Pro Met Arg Gly Asp Val Val Glu Thr Ile Ala Ser Ile Trp Arg | 980  |  | 985  |  | 990  |
| Glu Val Leu Arg Val Glu His Val Asp Tyr Arg Gln Asn Phe Phe Asp | 995  |  | 1000 |  | 1005 |
| Val Gly Gly His Ser Leu Met Leu Thr Arg Val Arg Gly Leu Leu Glu | 1010 |  | 1015 |  | 1020 |
| Glu Arg Leu Gly Leu Thr Leu Ser Val Val Asp Leu Phe Arg His Thr | 1025 |  | 1030 |  | 1035 |
| Thr Ile Glu Ser Leu Ala Gly Leu Ala Glu Lys Ser Glu Pro Ala Ala | 1045 |  | 1050 |  | 1055 |
| Ala Glu Pro Ala Ala Ala Val Ala Glu Asp Arg Ile Ala Val Ile Gly | 1060 |  | 1065 |  | 1070 |

Met Ala Gly Arg Phe Pro Gly Ala Arg Asn Val Glu Glu Phe Trp Arg  
 1075 1080 1085

Asn Leu Arg Asp Gly Val Asp Ser Ile Ala Arg Leu Ser Pro Glu Asp  
 1090 1095 1100

Leu Leu Ala Gly Gly Ile Ser Pro Glu Val Phe Gln Asp Pro Ser Tyr  
 1105 1110 1115 1120

Val Pro Ala Lys Gly Leu Leu Asp Gly Ile Glu Phe Phe Asp Ala Ala  
 1125 1130 1135

Phe Phe Gly Tyr Ser Pro Arg Glu Ala Glu Ile Met Asp Pro Gln His  
 1140 1145 1150

Arg Val Phe Leu Glu Cys Ala Trp Glu Ala Met Glu Asn Ala Gly Tyr  
 1155 1160 1165

Ala Ala Arg Ser Tyr Lys Gly Ser Ile Gly Val Phe Ala Gly Cys Gly  
 1170 1175 1180

Val Asn Thr Tyr Leu Leu Asn Asn Leu Ala Thr Ala Glu Pro Phe Asp  
 1185 1190 1195 1200

Phe Ser Arg Pro Ser Ala Tyr Gln Leu Leu Thr Ala Asn Asp Lys Asp  
 1205 1210 1215

Phe Leu Ala Thr Arg Val Ser Tyr Lys Leu Asn Leu Arg Gly Pro Ser  
 1220 1225 1230

Leu Thr Val Gln Thr Ala Cys Ser Thr Ser Leu Val Ser Val Val Met  
 1235 1240 1245

Ala Cys Glu Ser Leu Gln Arg Gly Ala Ser Asp Ile Ala Leu Ala Gly  
 1250 1255 1260

Gly Val Ala Ile Asn Val Pro Gln Ser Val Gly Tyr Leu His Gln Pro  
 1265 1270 1275 1280

Gly Met Ile Leu Ser Pro Asp Gly Arg Cys Arg Ala Phe Asp Glu Ser  
 1285 1290 1295

Ala Gln Gly Thr Val Pro Gly Asn Gly Ala Gly Val Val Val Leu Lys  
 1300 1305 1310

Arg Leu Ser Arg Ala Leu Ala Asp Gly Asp Thr Ile Tyr Ala Val Ile  
 1315 1320 1325

Arg Gly Ala Ala Ile Asn Asn Asp Gly Ala Glu Arg Met Gly Phe Thr  
 1330 1335 1340

Ala Pro Gly Val Asp Gly Gln Thr Arg Leu Ile Arg Arg Thr Gln Glu  
1345 1350 1355 1360

Met Ala Gly Val Lys Pro Glu Ser Ile Gly Tyr Ile Glu Ala His Gly  
1365 1370 1375

Thr Ala Thr Pro Leu Gly Asp Pro Val Glu Ile Ala Ala Ile Ala Ala  
1380 1385 1390

Asn Phe Pro Lys Asn Gly Ser Gly Asp Val Tyr Ile Gly Ser Val Lys  
1395 1400 1405

Thr Asn Ile Gly His Leu Asp Val Ala Ala Gly Val Ala Gly Leu Ile  
1410 1415 1420

Lys Thr Val Leu Ala Val His Arg Gly Gln Ile Pro Pro Ser Leu Asn  
1425 1430 1435 1440

Phe Gln Arg Pro Asn Pro Arg Ile Asp Phe Ala Asn Thr Pro Phe Arg  
1445 1450 1455

Val Ser Thr Arg Leu Leu Asp Trp Pro Ala Gly Lys Thr Pro Arg Arg  
1460 1465 1470

Ala Ala Val Ser Ser Phe Gly Ile Gly Gly Thr Asn Ala His Val Ile  
1475 1480 1485

Leu Glu Gln Ala Pro Pro Val Thr Pro Ala Ala Ala Ala Pro Glu Arg  
1490 1495 1500

Ser Ala His Val Leu Cys Leu Ser Ala Asn Thr Asp Ala Ala Leu Glu  
1505 1510 1515 1520

Glu Leu Val Arg Ser Tyr Arg Gly His Met Asp Asn Gln Pro Gly Leu  
1525 1530 1535

Ser Phe Gly Asp Val Ala Phe Thr Ala Asn Ala Gly Arg Val His Phe  
1540 1545 1550

Pro His Arg Ile Cys Ile Val Ala Arg Ser Ser Asp Glu Ala Arg Gln  
1555 1560 1565

Arg Leu Thr Glu Ala Arg Arg Val Arg Ile Ala Gln Thr Arg Pro Lys  
1570 1575 1580

Ile Ala Phe Leu Phe Thr Gly Gln Gly Ala Gln Tyr Ala Gly Met Gly  
1585 1590 1595 1600

Arg Gln Phe Tyr Glu Ser Gln Pro Val Phe Arg Ala Ala Met Asp Glu

| 1605                                                                                   | 1610 | 1615 |
|----------------------------------------------------------------------------------------|------|------|
| Cys Ala Ala Leu Leu Asn Gly Arg Leu Asp Leu Pro Ala Leu Leu Ala<br>1620 1625 1630      |      |      |
| Asp Asp Ala Leu Leu Asp Ala Thr Ala Gly Ala Gln Pro Ala Leu Phe<br>1635 1640 1645      |      |      |
| Ala Leu Gln Trp Ala Leu Ala Gln Leu Trp Lys Ser Trp Gly Val Thr<br>1650 1655 1660      |      |      |
| Pro Asp Leu Val Met Gly His Ser Val Gly Glu Tyr Ala Ala Ala Cys<br>1665 1670 1675 1680 |      |      |
| Ile Ala Gly Ala Val Ser Leu Pro Asp Ala Leu Gly Leu Val Ala Glu<br>1685 1690 1695      |      |      |
| Arg Gly Arg Leu Met Gln Asn Leu Pro Glu Gly Ala Met Ala Ala Val<br>1700 1705 1710      |      |      |
| Ser Ala Gly Glu Gln Arg Cys Ala Ala Ala Ile Thr Ser Arg Val Ser<br>1715 1720 1725      |      |      |
| Ile Ala Ala Ile Asn Gly Pro Ala Glu Val Val Ile Ser Gly Ala Pro<br>1730 1735 1740      |      |      |
| Gln Asp Ile Glu Ser Ala Leu Ala Thr Leu Arg Ala Glu Gly Ile Lys<br>1745 1750 1755 1760 |      |      |
| Thr Gln Met Leu Ala Val Ala Arg Ala Phe His Ser Ser Ser Met Asp<br>1765 1770 1775      |      |      |
| Pro Ile Leu Ala Asp Leu Gln Arg Arg Ala Ala Ala Ile Ala Trp Arg<br>1780 1785 1790      |      |      |
| Asn Pro Ser Ile Gly Leu Val Ser Asn Leu Thr Gly Lys Leu Ala Gly<br>1795 1800 1805      |      |      |
| Glu Gly Gln Leu Ala Asn Pro Leu Tyr Trp Arg Asp His Ala Arg Asn<br>1810 1815 1820      |      |      |
| Pro Val Arg Phe Ala Asp Gly Ile Gln Thr Leu Lys Asp Glu Gly Cys<br>1825 1830 1835 1840 |      |      |
| Asp Val Phe Leu Glu Ile Gly Pro Lys Pro Val Leu Leu Gly Met Gly<br>1845 1850 1855      |      |      |
| Gln Lys Cys Leu Pro Asp Asp Ala Lys Gln Trp Leu Pro Ser Leu Arg<br>1860 1865 1870      |      |      |

Lys Gly Arg Asp Glu Trp Glu Thr Ile Leu Ser Ser Val Ala Thr Leu  
 1875 1880 1885  
 Tyr Gln Gly Gly Phe Asp Ile Asp Trp Gln Glu Phe Asp Arg Pro Tyr  
 1890 1895 1900  
 Ser Arg Arg Arg Val Ala Leu Pro Ala Tyr Pro Phe Glu Arg Arg Arg  
 1905 1910 1915 1920  
 His Trp Ile Glu Arg Ser Ser Arg Pro Glu Pro Val Ala Val Ala Ser  
 1925 1930 1935  
 Gly Leu Val Gly Cys Arg Leu Ser Leu Pro Val Ala Asp Val Ile Phe  
 1940 1945 1950  
 Glu Ser Lys Leu Ser Thr Ala Ser Pro Leu Leu Ser Asp His Arg Tyr  
 1955 1960 1965  
 Tyr Gly Ser Val Val Ala Pro Ala Val Tyr Phe Leu Ala Met Ala Leu  
 1970 1975 1980  
 Glu Ala Ser Ala Glu Val Phe Gly Ala Gly Arg His Thr Leu Glu Asn  
 1985 1990 1995 2000  
 Val Asn Phe Ala His Pro Leu Ile Leu Ser Ala Glu Arg Asp Thr Ala  
 2005 2010 2015  
 Val Gln Leu Val Leu Ser Gln Ser Asp Asp Arg His Ala Ser Phe Arg  
 2020 2025 2030  
 Ile Leu Ser Leu Ser Asp Gly Ser Trp Asn Leu His Ala Ala Gly Asn  
 2035 2040 2045  
 Ile Ala Ala His Ala Gly Val Ala Pro Val Pro Arg Leu Val Asp Glu  
 2050 2055 2060  
 Arg Arg Pro Ala Val Asp Gly Asp Thr Tyr Tyr Ser Leu Leu Arg His  
 2065 2070 2075 2080  
 Leu Glu Ile Glu Leu Gly Pro Ser Tyr Arg Arg Ile Gln Arg Ile His  
 2085 2090 2095  
 Phe Gly Glu Gln Glu Ala Leu Ala Ala Ile Asp Ser Ala Thr Pro Leu  
 2100 2105 2110  
 Asn Pro Arg Cys Glu Leu Ala Glu Ala Gly Leu Gln Leu Leu Ser Ala  
 2115 2120 2125  
 Ala Ala Ser Pro Ala Leu Ala Asp Gly Ala Glu His Pro Ile Phe Ala  
 2130 2135 2140

Pro Leu Gly Ile Asp Arg Val Cys Phe Tyr Gly Ser Leu Glu Gly Ala  
 2145 2150 2155 2160  
 Val Trp Gly Ala Ala Gln Ile Leu Arg His Ser Pro Asp Gly Phe Thr  
 2165 2170 2175  
 Gly Glu Ala Gln Leu Leu Asp Ser Glu Gly Cys Val Leu Gly Glu Leu  
 2180 2185 2190  
 Gln Gly Val Ser Phe Arg Arg Val Thr Arg Ala Trp Ala Gln Arg Ser  
 2195 2200 2205  
 Glu Arg Lys Pro Glu Leu Tyr Glu Val Glu Trp Arg Pro Glu Pro Leu  
 2210 2215 2220  
 Arg Gln Pro Ser Arg Thr Leu Gln Pro Gly Ala Trp Leu Ile Leu Ala  
 2225 2230 2235 2240  
 Asp Ser Gly Gly Ala Ala Arg Ala Leu Ala Asp Ala Leu Thr Ala Gln  
 2245 2250 2255  
 Gly Glu Met Cys Val Thr Val Pro Pro Ala Gly Glu Tyr Met Ser Leu  
 2260 2265 2270  
 Val Gly Glu Arg Asp Trp Arg Gly Ile Val Asn Leu Tyr Ser Leu Asp  
 2275 2280 2285  
 Asp Tyr Glu Leu Gly Cys Arg Ser Thr Leu Ala Leu Val Lys Ser Leu  
 2290 2295 2300  
 Lys Ser Gly Pro Arg Leu Trp Leu Val Thr Ala Gly Ala Gln Ala Thr  
 2305 2310 2315 2320  
 Ser Ala Val His Asn Pro Met Gln Ala Ala Leu Trp Gly Phe Gly Arg  
 2325 2330 2335  
 Val Ile Ala Arg Glu His Pro Asp Leu Trp Gly Gly Leu Ile Asp Leu  
 2340 2345 2350  
 Asp Pro Asp Asp Ala His Ala Ser Ala Ala Gly Ala Ala Ala Gln Met  
 2355 2360 2365  
 Arg Asp Phe Asp Gly Glu Asp Gln Ser Ala Trp Arg Ser Asn Arg Arg  
 2370 2375 2380  
 Tyr Val Pro Arg Leu Thr Arg Arg Pro Ser Ala Arg Ala Ala Val Arg  
 2385 2390 2395 2400  
 Leu Val Ser Gly Ala Thr Tyr Leu Ile Thr Gly Gly Leu Gly Ala Leu

|                                                                 |      |      |
|-----------------------------------------------------------------|------|------|
| 2405                                                            | 2410 | 2415 |
| Gly Leu Thr Val Ala Lys Trp Met Val Glu His Gly Ala Thr Arg Val |      |      |
| 2420                                                            | 2425 | 2430 |
| Val Leu Ala Gly Arg Arg Pro Pro Asn Glu Glu Gln Gln Arg Val Leu |      |      |
| 2435                                                            | 2440 | 2445 |
| Gln Gln Ile Gly Ala Thr Ala Glu Thr Val Asp Val Ser Arg Glu Glu |      |      |
| 2450                                                            | 2455 | 2460 |
| Glu Val Ala Asp Leu Ile Arg Arg Ile His Thr Glu Thr Ser Pro Leu |      |      |
| 2465                                                            | 2470 | 2475 |
| Arg Gly Val Ile His Ala Ala Gly Val Leu Asp Asp Gly Val Leu Leu |      |      |
| 2485                                                            | 2490 | 2495 |
| Asn Gln Asp Trp Thr Arg Ile Ala Ser Val Met Ala Pro Lys Ala Glu |      |      |
| 2500                                                            | 2505 | 2510 |
| Gly Ala Val His Leu His His His Thr Arg Asp Leu Pro Leu Asp Phe |      |      |
| 2515                                                            | 2520 | 2525 |
| Phe Val Leu Phe Ser Ser Ala Ser Ser Leu Leu Gly Pro Ala Gly Gln |      |      |
| 2530                                                            | 2535 | 2540 |
| Ala Gly Tyr Ala Ala Ala Asn Ala Val Leu Asp Ala Leu Ala His His |      |      |
| 2545                                                            | 2550 | 2555 |
| Arg Arg Gly Leu Gly Leu Pro Ala Thr Ser Ile Asn Trp Gly Arg Trp |      |      |
| 2565                                                            | 2570 | 2575 |
| Ser Gly Ala Gly Met Ala Ala Arg Thr Ser Gln Ser Met Ala Gly Val |      |      |
| 2580                                                            | 2585 | 2590 |
| Ala Ser Leu Ser Val Asp Glu Gly Leu His Ile Leu Glu Ala Val Leu |      |      |
| 2595                                                            | 2600 | 2605 |
| His Glu Cys Pro Ile Gln Ile Ala Ala Leu Pro Ala Gly Ser Ile Thr |      |      |
| 2610                                                            | 2615 | 2620 |
| Gly Glu Leu Leu Arg Pro Ala Ala Leu Pro Ser Pro Gln Leu Arg Thr |      |      |
| 2625                                                            | 2630 | 2635 |
| Arg Leu Asn Glu Ala Thr Pro Arg Gln Arg Glu Ala Ile Leu Ile Ala |      |      |
| 2645                                                            | 2650 | 2655 |
| His Ile Arg Glu Ser Leu Ala Arg Phe Val Gly Ile Ala Thr Ser Thr |      |      |
| 2660                                                            | 2665 | 2670 |



Pro Leu Asp Pro Gln Gln Pro Leu Gly Glu Leu Gly Leu Asp Ser Leu  
 2675 2680 2685

Met Ala Ile Glu Leu Arg Asn Ser Leu Ser Gln Ser Leu Gly Gln Pro  
 2690 2695 2700

Leu Pro Ala Ser Leu Leu Phe Asp Tyr Pro Ser Leu Asp Ala Ile Val  
 2705 2710 2715 2720

Ser Tyr Val Leu His Ala Val Phe Pro Pro Glu Ala Ser Pro Val Glu  
 2725 2730 2735

Ala Pro Glu Phe Glu Asn Leu Ala Arg Glu Glu Leu Glu Ala Leu Leu  
 2740 2745 2750

Asp Ser Arg Leu Ala Gln Val Asp Gln Trp Leu Glu Thr Gln  
 2755 2760 2765

<210> 123

<211> 1763

<212> PRT

<213> bacterie

<400> 123

Met Ser Gly Ser Asp Asp Leu Ser Lys Leu Arg Arg Ala Val Ile Ala  
 1 5 10 15

Leu Asp Lys Val Gln Lys Arg Ile Asp Gln Leu Glu Ser Ala Arg Ser  
 20 25 30

Glu Pro Ile Ala Leu Ile Gly Ala Gly Cys Arg Phe Pro Gly Ala Ser  
 35 40 45

Asn Leu Asp Ala Tyr Trp Ser Leu Leu Arg Glu Gly Arg Ser Ala Val  
 50 55 60

Arg Glu Val Pro Pro Asp Arg Trp Asp Ile Asp Ala Tyr Tyr Asp Pro  
 65 70 75 80

Asp Pro Gly Ala Thr Gly Arg Met Tyr Thr Arg Tyr Gly Gly Phe Ile  
 85 90 95

Asp Gln Val Asp Arg Phe Asp Ala Arg Phe Phe Gly Ile Ala Pro Arg  
 100 105 110

Glu Ala Ile Ser Leu Asp Pro Gln Gln Arg Leu Leu Leu Glu Val Thr  
 115 120 125

Trp Glu Ala Ile Glu Asn Ala Gly Leu Pro Pro Asp Arg Leu Ala Gly  
 130 135 140  
 Ser Arg Thr Gly Val Phe Met Gly Ile Phe Ser Asn Asp Tyr Tyr Asn  
 145 150 155 160  
 Leu Gln Met Arg Gly Gly Asp Ala His Ile Asp Ala Tyr Thr Gly Thr  
 165 170 175  
 Gly Asn Thr Ala Ser Val Ala Ala Gly Arg Leu Ser Tyr Ile Leu Gly  
 180 185 190  
 Leu Gln Gly Pro Asn Met Ala Ile Asp Thr Ala Cys Ser Ser Ser Leu  
 195 200 205  
 Val Ala Val His Leu Ala Cys Gln Ser Leu Arg Ser Gly Glu Ser Asp  
 210 215 220  
 Leu Ala Leu Ala Gly Gly Val Asn Leu Ile Leu Ser Pro Asp Arg Thr  
 225 230 235 240  
 Ile Tyr Phe Cys Lys Leu Lys Ala Met Ala Ala Asp Gly Arg Cys Lys  
 245 250 255  
 Ala Phe Asp Ala Ala Ala Asp Gly Tyr Val Arg Gly Glu Gly Cys Gly  
 260 265 270  
 Val Val Val Leu Lys Arg Leu Ser Asp Ala Leu Arg Asp Arg Asp Pro  
 275 280 285  
 Val Met Ala Val Ile Arg Gly Thr Ala Ile Asn Gln Asp Gly Arg Ser  
 290 295 300  
 Asn Gly Leu Thr Ala Pro Asn Gly Pro Ala Gln Glu Ala Val Ile Arg  
 305 310 315 320  
 Gln Ala Val Gly Asp Ala Arg Leu Gln Thr Leu Asp Val Ser Tyr Val  
 325 330 335  
 Glu Ala His Gly Thr Gly Thr Pro Leu Gly Asp Pro Ile Glu Ala Gly  
 340 345 350  
 Ala Leu Ala Ala Ala Leu Gly Ala Gly Arg Thr Asn Gly Asn Lys Leu  
 355 360 365  
 Lys Leu Gly Ser Val Lys Thr Asn Phe Gly His Leu Glu Ala Ala Ala  
 370 375 380  
 Gly Val Ala Ala Leu Ile Lys Val Ala Leu Met Leu Gln Asn Glu Ala  
 385 390 395 400

Ile Pro Pro His Leu Asn Leu Thr Thr Pro Ser Pro His Ile Asp Trp  
 405 410 415  
 Asn Thr Leu Pro Leu Glu Ile Pro Ala Arg Leu Thr Pro Trp Pro Val  
 420 425 430  
 Ala Pro Gly Gly Arg Arg Val Ala Gly Ile Asn Ser Phe Gly Leu Ser  
 435 440 445  
 Gly Thr Asn Ala His Val Leu Ile Glu Gln Ala Pro Gln Gln Ala Ala  
 450 455 460  
 Ser Ser Thr Pro Ala Pro Tyr Leu Leu Pro Leu Ser Ala Arg Ser Pro  
 465 470 475 480  
 Glu Ala Leu Arg Asp Leu Ala Arg Ala Tyr Arg Asp Val Val Asn Asp  
 485 490 495  
 Asn Pro Ala Asp Thr Cys Tyr Thr Ala Cys Ala Arg Arg Thr Ser Tyr  
 500 505 510  
 Glu His Arg Ala Ala Phe Thr Gly Thr Asn Ala Gln Asp Leu Met Ala  
 515 520 525  
 Gly Leu Asp Ser Phe Leu Ala Gly Asn Pro Asn Arg Asp Thr Ala Thr  
 530 535 540  
 Gly Phe Val Pro Arg Gly Gln Lys Arg Lys Val Val Phe Val Leu Pro  
 545 550 555 560  
 Gly Gln Gly Ser Gln Trp Pro Gly Met Gly Arg Asp Leu Met Ala Ser  
 565 570 575  
 Glu Pro Val Phe Arg Ala Ala Ile Glu Glu Cys Gly Arg Ala Met Gln  
 580 585 590  
 Pro Tyr Val Asp Trp Ser Leu Thr Gln Glu Leu Gln Gly Pro Leu Asp  
 595 600 605  
 Arg Ile Asp Val Ile Gln Pro Ala Leu Phe Ala Val Gly Val Ala Leu  
 610 615 620  
 Ala Gly Leu Trp Arg His Trp Gly Ile Glu Pro Asp Ala Val Ile Gly  
 625 630 635 640  
 His Ser Met Gly Glu Val Ala Ala Ala His Ile Ala Gly Ala Leu Thr  
 645 650 655  
 Leu Asp Glu Ala Ala Arg Val Ile Cys Leu Arg Ser Arg Met Leu Ala

|                                                                                    |     |     |
|------------------------------------------------------------------------------------|-----|-----|
| 660                                                                                | 665 | 670 |
| Gly Val Arg Gly Gln Gly Glu Met Ala Val Val Glu Leu Ala Leu Asp<br>675 680 685     |     |     |
| Glu Ala Ile Ala Ala Ile Ala Gly Arg Ser Asp Arg Val Ser Ile Ala<br>690 695 700     |     |     |
| Ala Ser Asn Ser Pro Arg Ser Thr Val Leu Ser Gly Asp Ser Ala Ala<br>705 710 715 720 |     |     |
| Leu Gly Glu Leu Leu Arg Glu Leu Glu Ala Lys Asp Val Phe Cys Arg<br>725 730 735     |     |     |
| Arg Val Lys Val Asp Ile Ala Ser His Ser His Leu Met Asp Ser Val<br>740 745 750     |     |     |
| Cys Ala Ala Leu Pro Gly Val Val Gly Ala Leu Gln Pro Arg Pro Ala<br>755 760 765     |     |     |
| Ala Leu Gly Met Tyr Ser Thr Val Thr Gly Ala Ala Ile Ser Gly Glu<br>770 775 780     |     |     |
| Glu Leu Val Ser Ala Tyr Trp Ala Arg Asn Leu Arg Gln Pro Val Met<br>785 790 795 800 |     |     |
| Leu Ser Thr Ala Val Ala Ala Ala Ala Gly Gly His Asp Val Phe<br>805 810 815         |     |     |
| Leu Glu Leu Ser Pro His Pro Leu Leu Val Gln Pro Ile Gln Glu Thr<br>820 825 830     |     |     |
| Leu Gly Asp Arg Ala Ala Ile Ala Ala Ala Ser Leu Arg Arg Asp Glu<br>835 840 845     |     |     |
| Asp Gly Asn Leu Ala Leu Arg Arg Thr Leu Gly Ala Leu Leu Thr Asn<br>850 855 860     |     |     |
| Gly Val Thr Pro Asp Trp Ser Arg Ile Tyr Pro Asn Gly Gly Gln Thr<br>865 870 875 880 |     |     |
| Arg Arg Leu Pro Asn Tyr Pro Trp Gln Arg Glu Arg Tyr Trp Ile Asp<br>885 890 895     |     |     |
| Ile Arg Pro Pro Gln Val Glu Ser Gln Ala Leu Pro Gly Arg Arg Ile<br>900 905 910     |     |     |
| Pro Ser Pro Leu Pro Glu Met Gln Phe Glu Ser Thr Val Glu Ala Lys<br>915 920 925     |     |     |

Asp Phe Ala Asp His Arg Leu His Asp Val Ile Val Thr Pro Gly Ala  
 930 935 940  
 Trp His Leu Ala Met Ala Leu Ala Ala Ala Arg Gln Gly Leu Gly Ala  
 945 950 955 960  
 Gly Pro His His Val Glu His Val Ser Leu Thr Gly Ala Leu Thr Leu  
 965 970 975  
 Pro Glu Asn Asp Ala Ala Arg Gln Val Gln Leu Val Leu Arg His Glu  
 980 985 990  
 Glu Gly Gly Gly Ala Ser Phe Arg Ile Tyr Ser Arg Glu Asp Ser Trp  
 995 1000 1005  
 Lys Leu His Ser Glu Gly Met Leu Gln Ala Gly Asp Ser Thr Ala Ser  
 1010 1015 1020  
 Ile Asp Leu Asp Ala Ile Arg Ala Arg Cys Thr Ala Glu Leu Thr Ala  
 1025 1030 1035 1040  
 Asp Ala Phe Tyr Ser Arg Leu Trp Asp Arg Gly Tyr His Phe Gly Pro  
 1045 1050 1055  
 Thr Phe Arg Thr Ile Gly Pro Ile Trp Arg Gly Asn Gly Glu Val Leu  
 1060 1065 1070  
 Cys Arg Val Asp Ile Pro Leu Thr Glu Met Gln Thr Ile Asp Cys Cys  
 1075 1080 1085  
 Leu Gln Leu Pro Ala Ala Leu Val His His Asp Asp Leu Lys Asp Val  
 1090 1095 1100  
 His Val Pro Val Gly Leu Asp Arg Phe Ser Leu Ala Glu Val Pro Thr  
 1105 1110 1115 1120  
 Gly Pro Val Trp Gly Tyr Ala Val Leu Arg Pro Asp Ser Thr Val Asp  
 1125 1130 1135  
 Val Arg Leu Val Thr Gly Thr Gly Ser Val Val Ala Glu Leu Val Gly  
 1140 1145 1150  
 Leu Gln Ser Arg Val Ala His Ser Gly Gln Leu Gly Glu Ser Glu Ile  
 1155 1160 1165  
 Pro Thr Trp Thr Val Gln Trp Thr Ala Ser Val Arg Arg Gly Asp Ala  
 1170 1175 1180  
 Asn Ala Gly Asn Ala Gly Gly Pro Trp Leu Val Ile Gly Glu Pro Ala  
 1185 1190 1195 1200

Ile Ala Glu Thr Leu Gln Lys Arg Gly Gln Thr Cys Arg Thr Ala Asp  
 1205 1210 1215  
 Thr Cys Ser Gly Pro Pro Cys Arg Gln Ile Val Tyr Cys Pro Ser Pro  
 1220 1225 1230  
 Arg Ile Asp Asp Leu Leu Ser Val Leu Arg Ser Ile Val Gln Ala Gly  
 1235 1240 1245  
 Trp Pro Glu Pro Pro Arg Leu Trp Leu Leu Thr Arg Gly Ser Ala Ala  
 1250 1255 1260  
 Val Leu Asn Ser Asp Lys Asp Ile Asp Ile Arg Gln Ala Trp Leu His  
 1265 1270 1275 1280  
 Gly Ile Gly Arg Thr Ile Ala Tyr Glu His Pro Glu Leu Arg Cys Thr  
 1285 1290 1295  
 Leu Val Asp Leu Asp Ala His Ser Asn Asp Cys Gly His Leu Ala Thr  
 1300 1305 1310  
 Leu Met Leu Ser Asn Ile Ala Glu Asp Gln Val Ala Ile Arg Gln Gly  
 1315 1320 1325  
 Thr Val Trp Ala Pro Arg Leu Ser Leu His Lys Ile Pro Ser Ala Pro  
 1330 1335 1340  
 Asp Val Ala Phe Arg Ala Asp Ala Thr Tyr Leu Ile Thr Gly Gly Leu  
 1345 1350 1355 1360  
 Gly Gly Leu Gly Leu Gln Val Ala Gly Trp Leu Ala Ala Ala Gly Ala  
 1365 1370 1375  
 Arg His Leu Val Leu Leu Gly Arg Ser Glu Arg Pro Arg Pro Gln Leu  
 1380 1385 1390  
 Glu Gly Val Asn Val Lys Ile Ile His Ala Asp Val Ala Asp Arg Gln  
 1395 1400 1405  
 Gln Leu Ser Asp Ala Leu Ala Ile Ile Asp Arg Asp Met Pro Pro Leu  
 1410 1415 1420  
 Arg Gly Val Phe His Leu Ala Gly Thr Leu Ala Asp Gly Met Leu Leu  
 1425 1430 1435 1440  
 Asn Leu Thr Thr Glu Arg Phe Glu Ala Ala Met Ala Pro Lys Val Ala  
 1445 1450 1455  
 Gly Ala Trp Asn Leu His Glu Leu Thr Ala Gly Arg Pro Leu Asp His

| 1460                                                                                   | 1465 | 1470 |
|----------------------------------------------------------------------------------------|------|------|
| Phe Val Leu Phe Ser Ser Ala Ser Ala Thr Val Gly Ser Pro Gly Gln<br>1475 1480 1485      |      |      |
| Gly Asn Tyr Ala Ala Gly Asn Ser Phe Leu Asp Ala Leu Ala His Leu<br>1490 1495 1500      |      |      |
| Arg Arg Ala Gln Gly Leu Pro Ala Val Ser Ile Ala Trp Gly Pro Trp<br>1505 1510 1515 1520 |      |      |
| Thr Gln Val Gly Leu Ala Ala Gln Ala Asn Arg Gly Asp Arg Leu Ala<br>1525 1530 1535      |      |      |
| Ala Arg Gly Ile Ser Val Ile Gln Pro Gln Gln Gly Leu Arg Ala Leu<br>1540 1545 1550      |      |      |
| Tyr Lys Ala Leu Thr Gln Ile Arg Pro His Val Ala Val Met Asn Phe<br>1555 1560 1565      |      |      |
| Asp Ile Ala Gln Trp Leu Arg Tyr Tyr Pro Ser Ala Ala Ser Met Ser<br>1570 1575 1580      |      |      |
| Leu Leu Ala Gly Ile Ala Pro Ala Ala Ala Asp Thr Lys Pro Ala Ala<br>1585 1590 1595 1600 |      |      |
| Asp Met Arg Ser Glu Leu Leu Ala Val Pro Ala Gly Arg Gln Arg Arg<br>1605 1610 1615      |      |      |
| Ala Arg Leu Glu Thr Leu Leu Met His Glu Ala Gly His Val Leu Arg<br>1620 1625 1630      |      |      |
| Phe Asp Pro Ala Lys Leu Asp Gly Arg Ala Thr Leu Gly Asp Leu Gly<br>1635 1640 1645      |      |      |
| Phe Asp Ser Leu Met Ala Leu Glu Phe Arg Asn Arg Leu Glu Ala Gly<br>1650 1655 1660      |      |      |
| Leu Arg Val Lys Leu Ser Ala Thr Leu Ile Trp Arg Tyr Pro Thr Phe<br>1665 1670 1675 1680 |      |      |
| Ser Ala Leu Ala Gln His Leu Ala Asp Lys Leu Gly Leu Pro Leu Glu<br>1685 1690 1695      |      |      |
| Ser Met Ala Gly Asn Ala Glu Pro Ser Thr Val Ala Ala Val Ala Thr<br>1700 1705 1710      |      |      |
| Leu Ala Thr Val Gly Thr Ala Ala Gly Glu Asp Arg Ser Pro Ala Ala<br>1715 1720 1725      |      |      |

Ala Asp Asp Leu Asp Ala Val Ala Asn Gln Ile Ala Gly Leu Gly Asp  
 1730 1735 1740

Lys Glu Ile Glu Ala Leu Leu Lys Gln Lys Phe Ala His Phe Ser Gly  
 1745 1750 1755 1760

Ala Ser Glu

<210> 124

<211> 2153

<212> PRT

<213> bacterie.

<400> 124

Met Ser Ser Ile Ser Glu Arg Phe Pro Asn Leu Thr Pro Leu Gln Gln  
 1 5 10 15

Ala Tyr Leu Thr Leu Glu His Met Gln Arg Arg Leu Asp Ala Ala Glu  
 20 25 30

Arg Asp Ala Arg Glu Pro Ile Ala Ile Val Gly Leu Gly Cys Arg Phe  
 35 40 45

Pro Gly Gly Asp Gly Pro Asp Glu Phe Trp Gln Met Leu Arg Ser Gly  
 50 55 60

Val Asp Ala Ile Arg Glu Val Pro Pro Gly Arg Trp Asp Glu Glu Ser  
 65 70 75 80

Val Arg Arg Ile Leu Lys Ser Leu Asn Pro Ala Thr Pro Val Lys Ile  
 85 90 95

Gln Ala Gly Phe Leu Asp Ser Ile Asp Gly Phe Asp Asn Asp Phe Phe  
 100 105 110

Gly Ile Ser Pro Arg Glu Ala Val Ser Ile Asp Pro Gln Gln Arg Leu  
 115 120 125

Leu Leu Glu Val Ala Trp Glu Ala Leu Glu Asp Ala Gly Gln Thr Met  
 130 135 140

Glu Gly Leu Ser Gly Ser Arg Thr Gly Val Phe Val Gly Ile His Ser  
 145 150 155 160

Gln Ser Ser Asp Tyr Phe Trp Met Gln Thr Ala Asp Gly Ala Arg Ile  
 165 170 175



Asp Pro Tyr Thr Ala Thr Gly Thr Ala His Ser Val Ile Ala Gly Arg  
 180 185 190  
 Leu Ser Tyr Leu Leu Asn Leu Gln Gly Pro Ser Ile Ala Leu Asp Thr  
 195 200 205  
 Ala Cys Ser Ser Ser Leu Ala Ala Val His Leu Ala Cys Gln Ser Leu  
 210 215 220  
 Arg Ser Gly Glu Cys Thr Leu Ala Val Ala Gly Gly Val Asn Leu Arg  
 225 230 235 240  
 Phe Ser Pro Glu Phe Met Tyr Ala Thr Ser Lys Met Gly Thr Ala Ser  
 245 250 255  
 Pro Ser Gly Arg Cys Arg Ala Phe Asp Ala Ala Ala Asp Gly Ile Val  
 260 265 270  
 Phe Gly Glu Gly Cys Gly Val Val Val Leu Lys Arg Leu Ser Asp Ala  
 275 280 285  
 Leu Ala Ala Gly Asp Arg Val Trp Ala Val Val Arg Gly Ser Ala Val  
 290 295 300  
 Asn Gln Asp Gly Arg Ser Ala Gly Leu Thr Ala Pro Asn Val Val Ser  
 305 310 315 320  
 Gln Gln Val Val Ile Arg Ser Ala Leu Ala Asn Ala Gly Val Ala Ala  
 325 330 335  
 Gln Gln Ile Gly Tyr Ile Glu Ala His Gly Thr Gly Thr Pro Leu Gly  
 340 345 350  
 Asp Pro Ile Glu Ile Glu Ala Leu Ala Glu Thr Val Gly Leu Pro Arg  
 355 360 365  
 Pro Val Gly Asp Val Cys Ala Val Gly Ser Leu Lys Ser Asn Ile Gly  
 370 375 380  
 His Leu Glu Gly Ala Ala Gly Ile Ala Gly Leu Ile Lys Ala Val Leu  
 385 390 395 400  
 Ala Leu Ser His Glu Thr Ile Pro Pro Ser Leu His Val Arg Gln Leu  
 405 410 415  
 Asn Pro Asn Ile Arg Leu Glu Gly Thr Ser Leu Asp Ile Val Lys Glu  
 420 425 430  
 Val Arg Pro Trp Pro Ala Gly Ser Arg Arg Arg Phe Ala Gly Val Ser  
 435 440 445

Ala Phe Gly Trp Ser Gly Thr Asn Ala His Val Val Leu Glu Glu Ala  
450 455 460

Ala Pro Thr Gly Arg Gly Glu Ala Ala Ser Gly Phe His Ser Arg Pro  
465 470 475 480

Pro Ala Ala Ala Ala Arg Ala Ala Val Pro Leu Ala Glu Gly Asp Thr  
485 490 495

Gly Gly Thr Pro Asp Ile Ala Gly Thr Pro Asp Thr Ala Asp Thr Pro  
500 505 510

Asp Thr Ala Asp Thr Pro Asp Ile Ala Gly Thr Ala Gly Thr Ala Ala  
515 520 525

Thr Thr Gly Ile Ala Asp Ala Met Tyr Val Leu Pro Leu Ser Ala His  
530 535 540

Gly Ala Asp Glu Leu Arg Arg Val Ala Arg Ala Tyr Gly Glu Leu Leu  
545 550 555 560

Thr Ala Ser His Ala Pro Ser Leu Arg Asp Leu Cys Tyr Thr Ala Ala  
565 570 575

Val Arg Arg Thr His His Arg Cys Arg Leu Ala Val Ser Gly Arg Thr  
580 585 590

Ala Glu Glu Leu Ala Ala Gln Leu Gln Gly Ile Thr Ile Pro Ser Gln  
595 600 605

Arg Arg Lys Thr Val Phe Val Phe Ser Gly Gln Gly Ser Gln Trp Ile  
610 615 620

Gly Met Gly Arg Ser Trp Met Asp Arg Glu Pro Val Ile Arg Glu Ala  
625 630 635 640

Leu Glu Arg Cys Glu Ala Ala Met Arg Pro Tyr Val Asp Trp Ser Leu  
645 650 655

Lys Glu Glu Leu Ala Lys Leu Asp Arg Val Glu Val Ile Gln Pro Ala  
660 665 670

Leu Phe Ala Leu Gln Val Ala Ile Ala Ala Leu Trp Arg Ser Trp Gly  
675 680 685

Ile Glu Pro Asp Ala Val Ile Gly His Ser Met Gly Glu Val Ala Ala  
690 695 700

Ala His Val Ala Gly Ala Leu Thr Leu Gln Asp Ala Ala Arg Ile Ile

|                                                                 |     |     |  |     |  |     |
|-----------------------------------------------------------------|-----|-----|--|-----|--|-----|
| 705                                                             |     | 710 |  | 715 |  | 720 |
| Cys Ser Arg Ser Arg Leu Leu Ser Arg Ile Ser Gly Leu Gly Gly Met |     |     |  |     |  |     |
|                                                                 | 725 |     |  | 730 |  | 735 |
| Ala Met Val Glu Leu Pro Leu Ala Glu Cys Glu Ala Val Leu Ser Thr |     |     |  |     |  |     |
|                                                                 | 740 |     |  | 745 |  | 750 |
| Tyr Thr Glu Arg Leu Ser Pro Ala Val Ser Asn Gly Pro Asn Ser Thr |     |     |  |     |  |     |
|                                                                 | 755 |     |  | 760 |  | 765 |
| Val Ile Ser Gly Glu Val Glu Ala Leu Ala Glu Val Val Ala Thr Leu |     |     |  |     |  |     |
|                                                                 | 770 |     |  | 775 |  | 780 |
| Glu Arg Arg Gly Val Ser Cys Arg Pro Val Lys Val Asp Phe Ala Ala |     |     |  |     |  |     |
|                                                                 | 785 |     |  | 790 |  | 800 |
| His Ser Pro Gln Val Asp Pro Leu Cys Asp Glu Leu Leu Gln Ser Leu |     |     |  |     |  |     |
|                                                                 | 805 |     |  | 810 |  | 815 |
| Asp Gly Ile Gln Pro Arg Pro Ala Thr Ile Pro Phe Tyr Ser Thr Val |     |     |  |     |  |     |
|                                                                 | 820 |     |  | 825 |  | 830 |
| Thr Gly Ala Thr Leu Glu Thr Thr Ser Leu Asp Ser Thr Tyr Trp Ala |     |     |  |     |  |     |
|                                                                 | 835 |     |  | 840 |  | 845 |
| Arg Asn Leu Arg Ser Pro Val Leu Phe Trp Gln Gly Ile Arg His Leu |     |     |  |     |  |     |
|                                                                 | 850 |     |  | 855 |  | 860 |
| Ala Asp Ser Gly His Asp Val Phe Leu Glu Ile Ser Pro His Pro Ile |     |     |  |     |  |     |
|                                                                 | 865 |     |  | 870 |  | 875 |
| Leu Leu Pro Ala Ile Gly Gly Asn Ala Ala Leu Val Pro Ser Leu Arg |     |     |  |     |  |     |
|                                                                 | 885 |     |  | 890 |  | 895 |
| Arg Asp Gln Asp Glu Arg Gly Ser Met Leu Thr Ser Leu Gly Ala Leu |     |     |  |     |  |     |
|                                                                 | 900 |     |  | 905 |  | 910 |
| Tyr Glu Ala Gly His Thr Val Ala Trp Arg Thr Val Tyr Pro Ser Gly |     |     |  |     |  |     |
|                                                                 | 915 |     |  | 920 |  | 925 |
| Asn Cys Val Arg Leu Pro Arg Tyr Pro Trp Gln Arg Arg Arg Phe Trp |     |     |  |     |  |     |
|                                                                 | 930 |     |  | 935 |  | 940 |
| Leu Asp Ala Ser Pro Ala Arg His Ala Ile Thr Leu Gly Asn Pro Leu |     |     |  |     |  |     |
|                                                                 | 945 |     |  | 950 |  | 955 |
| Leu Gly Lys Arg Val Glu Ala Ser Thr Gln Pro Gly Thr Phe Phe Trp |     |     |  |     |  |     |
|                                                                 | 965 |     |  | 970 |  | 975 |

Glu Thr Glu Leu Ser Leu Ala Ser Val Pro Trp Leu Ala Asp His Arg  
 980 985 990  
 Val Gln Gly Glu Val Val Leu Pro Ala Thr Ala Tyr Leu Asp Met Ala  
 995 1000 1005  
 Leu Ala Gly Thr Ser Glu Thr Phe Gly Glu Ser Pro Cys Val Leu Glu  
 1010 1015 1020  
 His Val Thr Phe Thr Gln Met Leu Ile Val Pro Arg Asp Gly Ser Met  
 1025 1030 1035 1040  
 Thr Leu Gln Leu Ala Ile Ala Val Asp Arg Pro Gly Met Ala Ser Phe  
 1045 1050 1055  
 Arg Ile Ser Ser Arg Gln Ala Ser Thr Trp Val Leu His Ala Ser Gly  
 1060 1065 1070  
 Asp Ile Arg Gln Thr Pro Ala Asp Ala Ser Thr Val Pro Pro Asp Ser  
 1075 1080 1085  
 Ala Glu Thr Val Gln Ala Arg Cys Pro Thr Val Val Pro Ala Ala Glu  
 1090 1095 1100  
 Leu Trp Arg Gln Met Ala Glu His Gly Val Glu Tyr Gly Pro Ala Phe  
 1105 1110 1115 1120  
 Arg Ala Leu Glu Gln Ile Trp Ser Cys Pro Gly Glu Ala Ile Gly Arg  
 1125 1130 1135  
 Leu Arg Ser Ser Glu Thr Arg Ser Thr Ala Pro Ala Phe Leu Asp Ala  
 1140 1145 1150  
 Cys Leu Gln Ile Ile Ala Ala Ala Phe Gly Pro Ala Gly Gly Thr Trp  
 1155 1160 1165  
 Leu Pro Ala Gly Ile Asp Arg Met Arg Trp Leu His Pro Ala Arg Ser  
 1170 1175 1180  
 Val Val Trp Thr His Ala Arg Leu Glu Gly Pro Ile Ala Asp Leu Ser  
 1185 1190 1195 1200  
 Leu Leu Asp Gly Glu Gly Gln Leu Val Ala Arg Ile Glu Gly Leu Arg  
 1205 1210 1215  
 Leu Gln Arg Leu Asp Ala Ser Glu Arg Ile Asp Met Arg Gly Trp Leu  
 1220 1225 1230  
 His Glu Leu Arg Trp Val Ala Gln Pro His Ala Ala Ala Glu Pro Pro  
 1235 1240 1245

Ala Ala Arg Ala Ala Arg Ser Trp Leu Ile Val Gly Ala Val Asp Ser  
 1250 1255 1260

Ala Leu Thr Ala Trp Leu Arg Ala Thr Gly Asn Arg Val Thr Gln Thr  
 1265 1270 1275 1280

Ser Pro Glu Lys Leu Asp Glu Leu Gln Pro Pro Leu Glu Glu Ile Val  
 1285 1290 1295

Phe Leu Leu Glu His Glu Pro Ser Cys Asp Arg Ile Leu His Leu Leu  
 1300 1305 1310

Gln Thr Leu Gly Arg Thr Pro Trp Arg Gln Ala Pro Arg Leu Trp Leu  
 1315 1320 1325

Val Thr Arg Gly Ala Gln Pro Val Asp Gly Gln Ile Leu Gln Ala Gly  
 1330 1335 1340

Ile Ala Gln Ala Pro Phe Trp Gly Leu Gly Arg Thr Val His Tyr Glu  
 1345 1350 1355 1360

His Pro Glu Leu Asn Cys Thr Leu Ile Asp Leu Asp Pro Ala Gly Gly  
 1365 1370 1375

Glu Glu Glu Leu Leu His Glu Leu Leu Thr Asn Asn Gly Glu Asn Gln  
 1380 1385 1390

Ile Ala Phe Arg Gly Gly Ala Arg Tyr Val Ala Arg Val Ala Arg His  
 1395 1400 1405

Glu Ala Asp Met Gln Pro Ala Met Phe Lys Ala Gly Asp Arg Pro Phe  
 1410 1415 1420

Arg Leu Glu Ile Asp Ala Pro Gly Val Leu Asp Arg Leu Arg Leu Arg  
 1425 1430 1435 1440

Ala Thr Ser Arg Arg Pro Pro Gln Ala Gly Glu Val Glu Ile Glu Val  
 1445 1450 1455

Cys Ala Ala Gly Leu Asn Phe Leu Asp Val Leu Leu Ala Leu Gly Val  
 1460 1465 1470

Met Pro Asp Asp Ala Pro Gly Ala Ile Ala Gly Ser Pro Arg Leu Gly  
 1475 1480 1485

Gly Glu Cys Ser Gly Arg Ile Val Ala Met Gly Lys Gly Val Thr Asp  
 1490 1495 1500

Phe Arg Ile Gly Asp Glu Val Val Ala Leu Ala Pro Cys Ser Phe Gly

|                                                                 |      |      |      |
|-----------------------------------------------------------------|------|------|------|
| 1505                                                            | 1510 | 1515 | 1520 |
| Arg Phe Val Thr Thr Pro Ala Phe Arg Val Ala Leu Lys Pro Ala Asn |      |      |      |
| 1525                                                            |      | 1530 | 1535 |
| Ile Pro Ala Glu Gln Ala Ala Ala Leu Pro Ile Ala Phe Leu Thr Ala |      |      |      |
| 1540                                                            | 1545 |      | 1550 |
| Asp Tyr Ala Leu Ser Arg Ala Ala Arg Leu Ala Pro Gly Glu Arg Val |      |      |      |
| 1555                                                            | 1560 | 1565 |      |
| Leu Ile His Ala Ala Thr Gly Gly Val Gly Leu Ala Ala Ile Gln Ile |      |      |      |
| 1570                                                            | 1575 | 1580 |      |
| Ala Gln Arg Ala Gly Ala Glu Ile Phe Ala Thr Ala Gly Ser Pro Glu |      |      |      |
| 1585                                                            | 1590 | 1595 | 1600 |
| Lys Arg Ala Tyr Leu Arg Ser Leu Gly Ile Ala His Val Ser Asp Ser |      |      |      |
| 1605                                                            | 1610 |      | 1615 |
| Arg Ser Met Ala Phe Val Asp Asp Ile Arg Asn Trp Thr Asn Gln Glu |      |      |      |
| 1620                                                            | 1625 |      | 1630 |
| Gly Val Asp Val Val Leu Asn Ser Leu Ser Gly Asp Leu Leu Glu Ala |      |      |      |
| 1635                                                            | 1640 |      | 1645 |
| Ser Phe Asp Leu Leu Arg Asp His Gly Arg Phe Ile Glu Ile Gly Lys |      |      |      |
| 1650                                                            | 1655 |      | 1660 |
| Arg Asp Tyr Tyr Ala Gly Arg Lys Leu Gly Leu Arg Pro Phe Leu Lys |      |      |      |
| 1665                                                            | 1670 | 1675 | 1680 |
| Asn Leu Ser Tyr Thr Leu Val Asp Leu Leu Gly Met Ser Leu Lys Arg |      |      |      |
| 1685                                                            | 1690 |      | 1695 |
| Pro Ala Leu Thr Arg Glu Leu Leu Gln Glu Met Val Ala Lys Phe Glu |      |      |      |
| 1700                                                            | 1705 |      | 1710 |
| Ser Glu Thr Trp Arg Pro Leu Glu Thr Arg Val Thr Thr Ile Thr Glu |      |      |      |
| 1715                                                            | 1720 |      | 1725 |
| Ser Val Glu Ala Phe Arg Thr Met Ala Gln Ala Arg His Ile Gly Lys |      |      |      |
| 1730                                                            | 1735 |      | 1740 |
| Ile Val Met Ala Met Arg Asp Cys Ala Asn Ala Pro Ile Ala Pro Leu |      |      |      |
| 1745                                                            | 1750 | 1755 | 1760 |
| Arg Ser Ala Phe Asp Ser Glu Gly Thr Tyr Leu Ile Thr Gly Gly Leu |      |      |      |
| 1765                                                            | 1770 |      | 1775 |

Gly Gly Leu Gly Leu Thr Val Ala Arg Trp Met Ile Gly Arg Gly Ala  
 1780 1785 1790  
 Arg Arg Leu Val Leu Leu Ser Arg Arg Ala Pro Ser Pro Glu Val Gln  
 1795 1800 1805  
 Gln Ala Ile Ala Val Met Asp Ala Asp Val Arg Thr Val Gln Ala Asp  
 1810 1815 1820  
 Val Ser Gln Arg Asp Glu Leu Glu Arg Val Ile Ser Ser Ile Asp Arg  
 1825 1830 1835 1840  
 Leu Arg Gly Val Ile His Ala Ala Ala Val Leu Asp Asp Ala Leu Leu  
 1845 1850 1855  
 Leu Asn Gln Thr Glu Ala His Phe Arg Asn Val Met Ala Ala Lys Ile  
 1860 1865 1870  
 Asp Gly Ala Trp Asn Leu His Leu Leu Thr Arg Asp Cys Pro Leu Asp  
 1875 1880 1885  
 His Phe Val Leu Phe Ser Ser Ala Ala Gly Leu Leu Gly Ala Pro Ala  
 1890 1895 1900  
 Gln Gly Asn Tyr Ala Ala Ala Asn Ala Phe Leu Asp Ala Leu Ala Tyr  
 1905 1910 1915 1920  
 Tyr Arg Lys Ala Gln Gly Leu Pro Ala Leu Ser Ile Gly Trp Gly Ala  
 1925 1930 1935  
 Trp Ser Glu Val Gly Leu Ala Ala Ala Gln Asp Asn Arg Gly Ser Arg  
 1940 1945 1950  
 Leu Ala Leu Arg Gly Met Glu Asn Leu Thr Pro Gln His Gly Leu Ala  
 1955 1960 1965  
 Ile Leu Glu Gln Leu Leu Asn Ser Ser Ala Cys His Val Ala Ala Met  
 1970 1975 1980  
 Pro Ile Asn Val Arg Gln Trp Arg Gln Phe Tyr Pro Lys Ala Ala Gln  
 1985 1990 1995 2000  
 Ser Ala Leu Phe Glu Leu Leu His Asp Asp Ala Ala Ser Glu Ala Asp  
 2005 2010 2015  
 Ala Pro Asn Ala Leu Arg Ala Arg Leu Gln Ser Ala Glu Pro Gln Thr  
 2020 2025 2030  
 Arg Arg Thr Leu Leu Glu Glu His Leu Gln Gln Gln Leu Ala Arg Val  
 2035 2040 2045

Leu Arg Ile Asp Ser Gln Thr Ile Asp Pro Leu Arg Pro Leu Lys Glu  
2050 2055 2060

Leu Gly Phe Asp Ser Leu Met Ala Leu Glu Phe Arg Asn Arg Leu Glu  
2065 2070 2075 2080

Leu Thr Leu Gly Leu Thr Leu Pro Ala Thr Leu Ile Trp Gly His Pro  
2085 2090 2095

Thr Leu Ala Gly Leu Ala Pro His Leu Ala Ser Gln Met Gly Leu Pro  
2100 2105 2110

Leu Val Glu Ala Gln Ala Ala Ala Ala Glu Gly Asp Ser Arg Ala  
2115 2120 2125

Met Lys Thr Ala Leu Ser Gly Leu Asp Asp Met Ser Glu Glu Ala Ala  
2130 2135 2140

Val Ala Ala Leu Arg Gly Ala Arg Ser  
2145 2150

<210> 125

<211> 1695

<212> PRT

<213> bacterie

<400> 125

Met Arg Glu Lys Ile Ala Pro Met Ser Ser Val Lys Leu Ala Leu Leu  
1 5 10 15

Ala Arg Asn Met Arg Gln Asn Ile Ala Gly Phe Asp Leu Val His Ala  
20 25 30

Glu Pro Ile Ala Ile Val Gly Met Ala Cys Arg Phe Pro Gly Gly Ala  
35 40 45

Lys Asn Pro Asp Ala Phe Trp Thr Leu Leu Lys Asn Gly Val Asp Gly  
50 55 60

Val Thr Glu Val Pro Pro Asp Arg Trp Asn Ser Asp Gln Tyr Tyr Ser  
65 70 75 80

Ser Asp Pro Asp Ala Pro Gly Lys Ala Tyr Ala Arg Tyr Ala Ala Phe  
85 90 95

Leu Glu Arg Ile Asp Gly Phe Asp Ala Glu Phe Phe Gly Ile Ser Pro  
100 105 110



Arg Glu Ala Leu Asn Met Asp Pro Gln Gln Arg Leu Leu Leu Glu Val  
 115 120 125  
 Cys Trp Glu Ala Ala Glu Asp Ala Gly Ile Ser Pro Gly Pro Leu Ala  
 130 135 140  
 Gly Ser Ala Thr Gly Val Phe Ala Gly Ser Cys Ala Gln Asp Phe Gly  
 145 150 155 160  
 Leu Phe Gln Tyr Ala Asp Pro Ala Arg Ile Gly Ala Trp Ser Gly Ser  
 165 170 175  
 Gly Val Ala His Ser Met Leu Ala Asn Arg Ile Ser Tyr Leu Leu Asp  
 180 185 190  
 Leu Arg Gly Pro Ser Met Ala Val Asp Thr Ala Cys Ser Ser Ala Leu  
 195 200 205  
 Val Ala Val His Leu Ala Cys Gln Ser Leu Arg Arg Glu Cys Asp  
 210 215 220  
 Ala Ala Phe Ala Gly Gly Val Asn Leu Ile Leu Thr Pro Glu Gly Met  
 225 230 235 240  
 Ile Ala Leu Ser Lys Ala Arg Met Leu Ala Pro Asp Gly Arg Cys Lys  
 245 250 255  
 Thr Phe Asp Ala Ala Ala Asp Gly Tyr Val Arg Gly Glu Gly Cys Gly  
 260 265 270  
 Ile Val Leu Leu Lys Arg Leu Ser Asp Ala Leu Ala Asp Gly Asp Ala  
 275 280 285  
 Ile Arg Ala Val Ile Arg Gly Ser Ala Ile Asn Gln Asp Gly Arg Ser  
 290 295 300  
 Asn Gly Ile Thr Ala Pro Asn Leu Gln Ala Gln Lys Ala Val Leu Gln  
 305 310 315 320  
 Glu Ala Val Ala Asn Ala His Ile Asp Pro Ser His Val Ser Leu Ile  
 325 330 335  
 Glu Ala His Gly Thr Gly Thr Ser Leu Gly Asp Pro Ile Glu Ile Glu  
 340 345 350  
 Ala Leu Gln Ser Val Tyr Asp Ala Pro Asp Ser Ala Pro Cys Leu Leu  
 355 360 365  
 Gly Ser Val Lys Thr Asn Ile Gly His Leu Glu Gly Ala Ala Gly Ile

|                                                                 |                                             |     |     |     |
|-----------------------------------------------------------------|---------------------------------------------|-----|-----|-----|
| 370                                                             |                                             | 375 |     | 380 |
| Ala Gly Leu Ile Lys                                             | Ala Val Leu Ala Leu Gln His Arg Thr Ile Pro |     |     |     |
| 385                                                             | 390                                         |     | 395 | 400 |
| Pro His Leu His Phe Arg Arg Leu Asn Pro Asn Ile Ser Leu Asp Gly |                                             |     |     |     |
|                                                                 | 405                                         |     | 410 | 415 |
| Ser Arg Phe Arg Ile Ala Thr Glu Ser Ser Pro Trp Thr Ser Glu Gly |                                             |     |     |     |
|                                                                 | 420                                         |     | 425 | 430 |
| Arg Pro Arg Leu Ala Gly Val Ser Ser Phe Gly Phe Gly Gly Ser Asn |                                             |     |     |     |
|                                                                 | 435                                         |     | 440 | 445 |
| Ala His Val Ile Leu Glu Glu Ala Pro Ala Leu Pro Leu Pro Lys Pro |                                             |     |     |     |
|                                                                 | 450                                         |     | 455 | 460 |
| Val Thr Arg Pro Gln Leu Leu Thr Leu Ser Ala Arg Thr Asp Glu Ala |                                             |     |     |     |
|                                                                 | 465                                         |     | 470 | 475 |
| Leu Gly Glu Leu Ala Gly His Phe Ala Glu Phe Leu Gln Ser His Pro |                                             |     |     |     |
|                                                                 | 485                                         |     | 490 | 495 |
| Asn Ala Leu Leu Ser Asp Val Cys Phe Thr Ser Gln Val Gly Arg Asp |                                             |     |     |     |
|                                                                 | 500                                         |     | 505 | 510 |
| Ala Tyr Ser His Arg Leu Ala Ile Thr Ala Ala Asp Ala Ala Glu Ala |                                             |     |     |     |
|                                                                 | 515                                         |     | 520 | 525 |
| Val Ala Ala Leu Ala Ala Ala Pro Arg Arg Glu Val Ser Leu Arg Arg |                                             |     |     |     |
|                                                                 | 530                                         |     | 535 | 540 |
| Arg Pro Ala Ile Ala Phe Leu Phe Thr Gly Gln Gly Ala Gln Tyr Ala |                                             |     |     |     |
|                                                                 | 545                                         |     | 550 | 555 |
| Gly Met Gly Ala Glu Leu Tyr Lys Thr Gln Pro Val Phe Arg Asp Ala |                                             |     |     |     |
|                                                                 | 565                                         |     | 570 | 575 |
| Leu Asp Arg Cys Ala Asp Trp Leu Arg Pro Gln Leu Asp Val Pro Leu |                                             |     |     |     |
|                                                                 | 580                                         |     | 585 | 590 |
| Thr Val Leu Leu Phe Glu Ser Val Ser Pro Leu His Glu Thr Ala Tyr |                                             |     |     |     |
|                                                                 | 595                                         |     | 600 | 605 |
| Thr Gln Pro Ala Met Phe Ala Leu Glu Trp Ala Leu Ala Gln Phe Trp |                                             |     |     |     |
|                                                                 | 610                                         |     | 615 | 620 |
| Leu Ser Leu Gly Val Arg Pro Asp Tyr Val Leu Gly His Ser Leu Gly |                                             |     |     |     |
|                                                                 | 625                                         |     | 630 | 635 |
|                                                                 |                                             |     |     | 640 |

Glu Tyr Val Ala Ala Cys Val Ala Gly Ala Phe Ser Val Glu Asp Gly  
 645 650 655  
 Leu Arg Leu Val Thr Ala Arg Gly Arg Leu Val Asn Ala Leu Pro Arg  
 660 665 670  
 Gly Lys Ala Val Ile Val His Ala Asn Pro Ser Arg Ile Ala Ala Leu  
 675 680 685  
 Ala Ala Lys Val Ala Val Ala Ala Ser Asn Ala Pro Asp Arg Thr Val  
 690 695 700  
 Ile Ser Gly Thr Ala Ala Glu Ile Ala Glu Ala Gln Asp Asp Leu His  
 705 710 715 720  
 Arg Ala Gly Val Glu Thr Arg Glu Leu Asn Val Ser His Ala Phe His  
 725 730 735  
 Ser Pro Leu Met Asp Pro Ile Leu Asp Lys Phe Glu Ala Leu Ala Gly  
 740 745 750  
 Ala Ile Ala Tyr Gln Pro Leu Ala Ile Pro Leu Val Ser Asn Val Ser  
 755 760 765  
 Gly Ala Val Leu Pro Lys Gly Thr Thr Leu Asp Ala Arg Tyr Trp Arg  
 770 775 780  
 Arg Gln Leu Arg Glu Thr Val Gln Phe Glu Ser Ala Met Arg Thr Leu  
 785 790 795 800  
 Ala Asp Arg Glu Cys Lys Leu Phe Leu Glu Ile Gly Pro His Pro Thr  
 805 810 815  
 Leu Thr Thr Leu Gly Arg Tyr Cys Leu Pro Asp Asp Gly Ala Val Trp  
 820 825 830  
 Leu His Ser Leu Ser Lys Gly Arg Ser Asp Trp Ser Val Leu Leu Glu  
 835 840 845  
 Ser Leu Gly Gly Leu Phe Thr Ala Gly Val Asn Pro Asp Trp Arg Gly  
 850 855 860  
 Leu Tyr Ala Gly Glu Ser Pro Ser Arg Val Ala Leu Pro Thr Tyr Pro  
 865 870 875 880  
 Phe Gln Arg Asp Thr Phe Ser Leu Arg Arg Val Pro Ala Arg Glu Pro  
 885 890 895  
 Ala Arg Gly Gly Met Leu Gly Ala Arg Leu Asn Ser Ala Leu Gly Asp  
 900 905 910

Val Ile Phe Glu Asn Ser Leu Thr Thr Glu Thr Pro Leu Leu His Glu  
 915 920 925  
 His Val Ile Tyr Asp Ala Val Ile Val Pro Gly Ala Trp His Val Ser  
 930 935 940  
 Ala Phe Leu Glu Ala Ala Gln Glu Val Phe Gly Pro Val Pro Cys Ala  
 945 950 955 960  
 Val Ser Asp Val Met Met Arg Gln Ala Leu Ala Ile Pro Pro Asp Thr  
 965 970 975  
 Pro Val Thr Val Gln Ala Ile Val Thr Pro Gly Glu Asp Gly Glu Ala  
 980 985 990  
 Lys Val Gln Val Phe Ser Gln Asp Gly Asp Ser Trp Lys Leu His Thr  
 995 1000 1005  
 Ala Ala Ser Leu Arg Ala Ala Thr Ala Gly Ala Val His Phe Glu Leu  
 1010 1015 1020  
 Pro Ala Gln Pro Ser Glu Val Ile Ser Gly Asp Ala Phe Tyr Gly Ala  
 1025 1030 1035 1040  
 Met Asn Ala Arg Gly Val Asp Leu Gly Pro Ala Phe Ser Trp Val Glu  
 1045 1050 1055  
 Glu Val Trp Arg Arg Asp Gly Glu Ala Leu Gly Arg Met Arg Leu Pro  
 1060 1065 1070  
 Val Ala Glu Asp Gly Ala Asn Ala Tyr Arg Leu His Pro Gly Leu Ile  
 1075 1080 1085  
 Asp Ser Cys Phe Gln Val Phe Gly Ala Thr Trp Pro Ala Glu Arg Cys  
 1090 1095 1100  
 Gln Pro Gly Ala Tyr Val Pro Val Gly Ile Glu Ala Val Arg Phe Tyr  
 1105 1110 1115 1120  
 Arg Pro Pro Ala Gly Ser Leu Arg Cys His Ala Arg Leu Arg Pro Ser  
 1125 1130 1135  
 Ser Ser Gly Pro Phe Val Gly Asp Leu Thr Leu Val Glu Glu Thr Gly  
 1140 1145 1150  
 Ala Val Ile Ala Glu Phe Ser Gly Leu Ala Val Met His Ala Gly Thr  
 1155 1160 1165  
 Leu Gln Ser Ala Gln Ser Trp Leu Gln Asp Val Gln Trp Gln Glu Cys

| 1170                                                            | 1175 | 1180      |
|-----------------------------------------------------------------|------|-----------|
| Glu Arg Ser Thr Thr Leu Lys Ser Asp Gly Pro Gly Lys Pro Glu Asp |      |           |
| 1185                                                            | 1190 | 1195 1200 |
| Trp Leu Leu Cys Ala Gly Ala Asp Asp Val Ala Gly Leu Met Pro Gln |      |           |
| 1205                                                            | 1210 | 1215      |
| Glu Leu Arg Val Val Ser Gly Val Thr Leu Arg Gln Ala Leu Glu Gln |      |           |
| 1220                                                            | 1225 | 1230      |
| Thr Gln Thr Leu Val Gly Arg Pro Ala Arg Leu Trp Leu Ile Thr Arg |      |           |
| 1235                                                            | 1240 | 1245      |
| Gly Val His Arg Ile Ser Asp Asp Asp Ala Thr Pro Val Asp Pro Phe |      |           |
| 1250                                                            | 1255 | 1260      |
| Gln Ala Pro Leu Trp Gly Leu Gly Gln Ala Ile Ala Arg Glu His Pro |      |           |
| 1265                                                            | 1270 | 1275 1280 |
| Glu Leu Trp Gly Gly Leu Ile Asp Leu Gly Cys Asp Asn Ala Asp Ile |      |           |
| 1285                                                            | 1290 | 1295      |
| Ala Ala Ala Met Leu Leu Asp Glu Ile Arg Tyr Ala Gly Asp Asp Lys |      |           |
| 1300                                                            | 1305 | 1310      |
| Ala Ile Ala Leu Arg Asn Gly Arg Arg Tyr Val Arg Arg Leu Val Arg |      |           |
| 1315                                                            | 1320 | 1325      |
| His Lys Glu Thr Ser Lys Arg Pro Pro Ala Ile Ser Ala Asp Gly Val |      |           |
| 1330                                                            | 1335 | 1340      |
| Tyr Leu Ile Thr Gly Gly Leu Gly Ala Leu Gly Arg Arg Val Ala Arg |      |           |
| 1345                                                            | 1350 | 1355 1360 |
| Arg Leu Ile Glu Gln Gly Ala Arg Arg Leu Val Leu Val Gly Arg His |      |           |
| 1365                                                            | 1370 | 1375      |
| Thr Glu Ala Val Ala Asp Leu Glu Gln Leu Gly Ala Ala Val Met Val |      |           |
| 1380                                                            | 1385 | 1390      |
| Ala Ala Cys Asp Val Ser Ser Glu Gln Gln Leu Ala Ala Leu Leu Ala |      |           |
| 1395                                                            | 1400 | 1405      |
| Asp Pro Arg Thr Gln Pro Leu Arg Gly Val Val His Ala Ala Gly Val |      |           |
| 1410                                                            | 1415 | 1420      |
| Leu Asp Asp Gly Val Val Thr Glu Gln Thr Trp Ala Arg Phe Glu Lys |      |           |
| 1425                                                            | 1430 | 1435 1440 |

Val Leu Ala Pro Lys Leu Gln Gly Ala Trp Asn Leu His Gln Leu Thr  
 1445 1450 1455  
 Arg His His Ala Leu Asp Phe Phe Val Leu Phe Ser Ser Ala Ala Ser  
 1460 1465 1470  
 Leu Leu Gly Ser Ala Gly Gln Ser Asn Tyr Ser Ala Ala Asn Ala Phe  
 1475 1480 1485  
 Leu Asp Ser Leu Ala His Met Arg Arg Ala Gln Gly Leu Pro Ala Leu  
 1490 1495 1500  
 Ser Ile Asn Trp Gly Pro Trp Ala Gly Glu Gly Met Ala Ala Arg Ile  
 1505 1510 1515 1520  
 Ala Arg Gln Gly Leu Pro Gly Val Pro Leu Leu Pro Pro Glu Val Gly  
 1525 1530 1535  
 Ala Arg Ile Phe Gly Asp Leu Leu Gly Glu Thr Ala Ala Gln Ile Ala  
 1540 1545 1550  
 Val Phe Gln Val Ser Ala Glu Lys Arg Arg Ser Pro Ala Ser Asp Pro  
 1555 1560 1565  
 Gly Phe Ile Gln Gln Leu Thr Glu Ala Ala Pro Glu Arg Arg Gln Glu  
 1570 1575 1580  
 Leu Leu Gln Met Arg Ile Arg Lys Gln Ala Gly Gly Val Leu Ala Leu  
 1585 1590 1595 1600  
 Asp Ala Ser Lys Thr Leu Asp Pro Arg Arg Pro Leu Lys Glu Tyr Gly  
 1605 1610 1615  
 Leu Asp Ser Leu Met Ala Leu Asp Leu Ala Arg Ala Ile Gly Glu Leu  
 1620 1625 1630  
 Val Arg Lys Ser Leu Pro Ala Thr Leu Leu Tyr Asp His Pro Thr Val  
 1635 1640 1645  
 Glu Lys Leu Ala Gly His Val Leu Arg Glu Leu Gly Leu Asp Val Pro  
 1650 1655 1660  
 Ser Asp Ser Leu Val Asp Glu Val Arg Gln Leu Ser Glu Gln Glu Met  
 1665 1670 1675 1680  
 Ala Ala Phe Ile Thr Glu Thr Leu His His Leu Gly Glu Glu Arg  
 1685 1690 1695

&lt;210&gt; 126

&lt;211&gt; 1434

&lt;212&gt; PRT

&lt;213&gt; bacterie

&lt;400&gt; 126

Met Ser Asp Leu Thr Pro Leu Gln Gln Ala Val Leu Ala Leu Lys Arg  
 1 5 10 15

Thr Arg Ala Arg Leu Asp Glu Leu Glu Ser Val His Asn Glu Pro Ile  
 20 25 30

Ala Ile Val Gly Met Ala Cys Arg Phe Pro Gly Ala Asp Ser Pro Glu  
 35 40 45

Ala Phe Trp Gln Leu Leu His Asp Gly Ile Asp Ala Ile Arg Glu Ile  
 50 55 60

Pro Ala Gly Arg Trp Asp Ala Asp Ala Phe Tyr Asp Pro Asp Pro Asn  
 65 70 75 80

Ala Pro Gly Lys Met Tyr Thr Arg Leu Gly Gly Phe Leu Asp Gly Ala  
 85 90 95

Val Asp Gly Phe Asp Ala Gly Phe Phe Gly Ile Thr Pro Arg Glu Val  
 100 105 110

Ala Gly Leu Asp Pro Gln Gln Arg Leu Leu Leu Glu Val Ala Trp Glu  
 115 120 125

Ala Leu Glu Arg Ala Gly Arg Pro Pro Asp Ser Leu Ala Gly Ser Asp  
 130 135 140

Thr Gly Val Phe Ile Gly Ile Ser Thr Asp Asp Tyr Ser Arg Leu Lys  
 145 150 155 160

Pro Thr Asp Pro Ala Leu Ile Asp Ala Tyr Thr Gly Thr Gly Thr Ala  
 165 170 175

Phe Ser Thr Ala Ala Gly Arg Ile Ser Tyr Leu Leu Gly Leu Gln Gly  
 180 185 190

Pro Asn Phe Pro Val Asp Thr Ala Cys Ser Ser Ser Leu Val Ala Val  
 195 200 205

His Leu Ala Cys Arg Ser Leu Gln Ser Arg Glu Cys Ser Met Ala Leu  
 210 215 220

Ala Gly Gly Val Asn Leu Ile Leu Ala Pro Glu Ser Thr Ile Tyr Phe  
 225 230 235 240

Cys Arg Leu Arg Ala Met Ala Ala Asp Gly Arg Cys Lys Ser Phe Ala  
 245 250 255  
 Ala Ser Ala Asp Gly Tyr Gly Arg Gly Glu Gly Cys Gly Met Leu Val  
 260 265 270  
 Leu Lys Arg Leu Ser Asp Ala Thr Arg Asp Gly Asp Arg Ile Leu Ala  
 275 280 285  
 Leu Ile Arg Gly Ser Ala Val Asn His Gly Gly Arg Ser Asn Gly Leu  
 290 295 300  
 Thr Ala Pro Asn Gly Pro Ala Gln Glu Ala Val Ile Arg Ala Ala Leu  
 305 310 315 320  
 Lys Asn Ala Gly Met Ala Pro Ala Asp Val Asp Tyr Val Glu Ala His  
 325 330 335  
 Gly Thr Gly Thr Pro Leu Gly Asp Pro Ile Glu Leu Arg Ala Met Ala  
 340 345 350  
 Ala Val Leu Gly Glu Gly Arg Ala Val Asp Ser Pro Leu Ile Val Gly  
 355 360 365  
 Ser Val Lys Thr Asn Phe Gly His Leu Glu Ala Ala Ala Gly Ile Ala  
 370 375 380  
 Gly Leu Ile Lys Thr Ile Leu Ala Leu Gln His Arg Glu Ile Pro Pro  
 385 390 395 400  
 His Leu His Phe Asn Ala Pro Asn Pro His Val Leu Trp Asn Glu Leu  
 405 410 415  
 Pro Leu Lys Ile Ala Thr Ala Cys Ser Pro Trp Pro Ser Asn Gly Arg  
 420 425 430  
 Pro Arg Val Ala Gly Val Ser Ser Phe Gly Ile Ser Gly Thr Asn Ser  
 435 440 445  
 His Val Val Leu Ala Glu Ala Lys Thr Asn Val Glu Ala Lys Thr Asn  
 450 455 460  
 Val Glu Ala Lys Thr Asn Val Glu Ala Lys Thr Ser Glu Glu Val Lys  
 465 470 475 480  
 Ala Ser Val Glu Ala Lys Gly Asn Val Glu Ala Lys Ala Ser Ala Ser  
 485 490 495  
 Val Pro Leu Leu Glu Gly Asp Ser Arg Pro Arg Ser Gly Gly Gly Gly



| 500                                                                                | 505 | 510 |
|------------------------------------------------------------------------------------|-----|-----|
| Ser Gly Arg Pro Pro Ser Arg Glu Glu Val Pro Val Pro Asp Gln Leu<br>515 520 525     |     |     |
| His Ala Glu Asp Gly Arg Glu Tyr Leu Leu Pro Leu Ser Ala Arg His<br>530 535 540     |     |     |
| Pro Gln Ala Leu Arg Asp Leu Ala Gly Ala Tyr Arg Asp Gly Arg Phe<br>545 550 555 560 |     |     |
| His Ala Pro Leu Ser Ala Leu Cys Ser Ala Ala Ser Leu Thr Arg Ser<br>565 570 575     |     |     |
| His Tyr Glu His Arg Ala Ala Phe Val Ala Ser Ser Leu Pro Glu Phe<br>580 585 590     |     |     |
| Asn Gln Leu Leu Glu Ala Phe Arg Arg Asn Glu Thr Asn Arg Gly Val<br>595 600 605     |     |     |
| Ala Thr Gly Phe Ala Asp Pro Gly Val Arg Pro Lys Leu Ala Phe Ile<br>610 615 620     |     |     |
| Phe Ser Gly Gln Gly Gly Gln Tyr Pro Arg Met Ala Tyr Arg Leu Tyr<br>625 630 635 640 |     |     |
| Ser Asp Glu Pro Val Phe Arg Ser Ala Ile Glu Arg Cys Asp Ala Ala<br>645 650 655     |     |     |
| Phe Arg Ser Phe Val Glu Trp Arg Leu Ala Asp Leu Leu Ala Asp Glu<br>660 665 670     |     |     |
| Ser Gly Ala Trp Leu Ser Gln Ile Asp Arg Val Gln Pro Ala Leu Phe<br>675 680 685     |     |     |
| Ala Val Gln Ile Ala Leu Val Glu Leu Leu Gln Ser Trp Gly Ile Arg<br>690 695 700     |     |     |
| Pro Asp Gly Val Ala Gly His Ser Met Gly Glu Val Ala Ala Ala His<br>705 710 715 720 |     |     |
| Val Ala Gly Ile Leu Thr Leu Glu Asp Ala Ala Arg Ile Ile Cys Arg<br>725 730 735     |     |     |
| Arg Ser Arg Leu Leu Leu Gly Leu Arg Gly Arg Gly Ala Met Ala Leu<br>740 745 750     |     |     |
| Val Glu Leu Pro Leu Asp Arg Ala Lys Ala Val Leu Ala Glu Arg Gly<br>755 760 765     |     |     |

Leu Thr Thr Val Ser Val Ala Ala Ser Asn Gly Pro Arg Ser Thr Val  
 770 775 780  
 Phe Ser Gly Asp Arg Val Ala Leu Glu His Leu Lys Asp Asp Phe Glu  
 785 790 795 800  
 Arg Arg Gly Val Phe Cys Arg Leu Ile Gln Val Asp Val Ala Ser His  
 805 810 815  
 Ser Ser Gln Val Asp Pro Leu Glu Asn Glu Leu Arg Gln Glu Leu Gly  
 820 825 830  
 Arg Val Ile Ala Lys Arg Ser Ala Val Pro Phe Phe Ser Thr Val Glu  
 835 840 845  
 Gly Gln Leu Ser Thr Gly Glu Ala Cys Asp Ala Ser Tyr Trp Val Ala  
 850 855 860  
 Asn Leu Arg Gln Pro Val Arg Phe Trp Glu Ser Leu Gln Ala Met Ala  
 865 870 875 880  
 Gly Asp Glu Phe Thr Gln Phe Leu Glu Ile Ser Pro His Pro Val Leu  
 885 890 895  
 Thr Pro Ser Ile Glu Asp Ser Leu Arg Thr Leu Gly Ile Asn Gly Leu  
 900 905 910  
 Val Arg Pro Val Leu Arg Arg Asp Glu Pro Glu Arg Arg Glu Leu Leu  
 915 920 925  
 Glu Leu Leu Ala Ala Leu Tyr Val Asn Gly Gln Arg Pro Asp Trp Arg  
 930 935 940  
 Ala Leu Ala Ser Ser Pro Asp Thr Arg Leu Asp Leu Pro Thr Tyr Pro  
 945 950 955 960  
 Trp Gln Arg Glu Arg Phe Trp Phe Ala Thr Ser Thr Arg Arg Ser Leu  
 965 970 975  
 Pro Ala Val Gly Gly His Pro Leu Leu Gly Arg Lys Val Glu Ile Ala  
 980 985 990  
 Leu Ala Pro Asp Thr His Val Trp Glu Ser Val Leu Ser Leu Asp Ala  
 995 1000 1005  
 Leu Pro Phe Leu Ala Asp His Arg Leu Asn Glu Leu Val Val Leu Pro  
 1010 1015 1020  
 Gly Ala Ala Tyr Val Glu Met Ala Leu Ala Ala Lys Glu Val Phe  
 1025 1030 1035 1040

Ala Gly Gly Cys Ser Leu Glu Glu Ile Arg Phe Glu Gln Met Leu Val  
1045 1050 1055

Val Pro Ser Ala Gly Ala Ser Arg Val Gln Val Ile Leu Glu Gly His  
1060 1065 1070

Ala Phe Arg Ile Ser Ser Leu Ala Glu Gly Gly Ser Asp Trp Thr Glu  
1075 1080 1085

His Ala Arg Gly Thr Met Ala Ala Ala Pro Asp Lys Val Ala Pro Thr  
1090 1095 1100

Val Ser Leu Pro Thr Leu Gly Asp Arg Ile Glu Gly Asp Asp Phe Tyr  
1105 1110 1115 1120

Ala Ala Phe Ala Ser Gln Gly Met His Tyr Gly Asp Thr Phe Arg Gly  
1125 1130 1135

Ile Ala Glu Val Trp Arg Asp Gly Glu Ala Val Ala Arg Leu Ser  
1140 1145 1150

Val Pro Asp Ala Val Arg Glu Ala Glu Ser Gly Tyr Thr Leu His Pro  
1155 1160 1165

Ala Leu Leu Asp Ala Cys Leu Gln Val Leu Gly Ala Thr Leu Gly Gly  
1170 1175 1180

Glu Gly Ser Ala Gly Pro Cys Val Pro Val Ala Ile Glu Arg Leu His  
1185 1190 1195 1200

Cys Phe Gly Arg Pro Ala Gly Asp Leu Arg Val His Ala Arg Leu Thr  
1205 1210 1215

Gly Arg Leu Glu Gly Asp Val Thr Leu Cys Asp Ala Glu Gly His Val  
1220 1225 1230

Ile Leu Glu Val Gln Gly Leu Arg Ala Gln Glu Leu Glu Arg Gln Ser  
1235 1240 1245

Glu Trp Phe His Ala Met Glu Trp Glu Pro Gln Leu Leu Ala Glu Ser  
1250 1255 1260

Pro Thr Ala Thr Val Ser Gly Ala Trp Leu Val Ile Ala Asp Ala Gly  
1265 1270 1275 1280

Gly Ile Ala Ala Ala Val Ala Arg Gly Leu Gly Thr Asn Thr Val Val  
1285 1290 1295

Ile Ser Gly Arg Asp Ala Glu Ile Pro Asp Gln Pro Tyr Arg Gly Val

| 1300                                                                    | 1305 | 1310      |
|-------------------------------------------------------------------------|------|-----------|
| Ile His Cys Gly Ser Leu Asp Glu Thr Glu Asp Glu Thr Asp Pro Ser<br>1315 | 1320 | 1325      |
| Ala Ala Gly Gly Thr Ala Cys Glu Asp Ile Leu Arg Ile Val Gln Glu<br>1330 | 1335 | 1340      |
| Phe Gly Val Gly Arg Ile Gln Leu Thr Lys Gln Ala Ser Asp Ala Glu<br>1345 | 1350 | 1355 1360 |
| Ser Gln His Pro Arg Ile Trp Leu Ile Thr Ala Gly Val His Ala Glu<br>1365 | 1370 | 1375      |
| His Leu Gln Met Pro Val Val Pro Ala Arg Ala Pro Val Trp Gly Leu<br>1380 | 1385 | 1390      |
| Gly Arg Thr Ile Ala Ala Glu His Pro Glu Phe Ala Cys Thr Cys Ile<br>1395 | 1400 | 1405      |
| Asp Leu Asp Thr Ala Gly Glu Val Glu Val Gln Ala Leu Cys Arg Glu<br>1410 | 1415 | 1420      |
| Ile Leu Ala Gly Ser Ser Glu Arg Gln Gly<br>1425                         | 1430 |           |